地理信息系统原理与技术

华一新　吴升　赵军喜　编著



DILIXINXIXITONGYUANLIYUJISHU

解放军出版社

地理信息系统
原理与技术

华一新　吴　升　赵军喜　编著

解放军出版社

图书在版编目(CIP)数据

地理信息系统原理与技术/华一新等著. —北京:
解放军出版社, 2000
ISBN 7-5065-3967-5

Ⅰ. 地… Ⅱ. 华… Ⅲ. 地理信息系统-教材
Ⅳ. P91

解放军出版社出版
(北京地安门西大街40号 邮政编码: 100035)
一二〇一工厂印刷 解放军出版社发行部发行
2001年2月第1版 2001年2月第1次印刷
开本: 850×1168毫米 1/32 印张: 15
字数: 406.6千字 印数: 1~3100册

序

在地理学和地图学的百年发展中，地理信息系统(GIS)的出现和实用化是一件大事。“地理系统研究人类赖以生存、生活和影响所及的整个自然环境与社会经济环境，并把地理环境看作是一个运动着的发生和发展的世界”(陈述彭，1991)。一旦对地观测技术和空间分析方法提供了技术与方法的保证，地理信息系统的出现就是必然的了。由于地理信息系统的数据处理和制图的特殊功能，使科学家、研究人员和管理人员第一次尝到了随心所欲地处理空间数据和充分表现自己创造性空间思维的甜头，通过地理空间图形的分析与制图来不断修正、完善和深化自己的认识。

在信息技术高速发展的今天，我们可以从不同的角度进入地理信息学的领域，因为它对硬件和软件都有很强的依赖性，但 GIS 最重要的技术与方法学的基础则是地图学。这首先是因为，制图是一种认识方法，是一个创造性的认知环境和发现客观规律的过程。例如《肿瘤地图集》出版后，很多医学专家开始重视从区域性规律和环境诱因上去做进一步探索；魏格纳(A. Wegener 1880～1930)的板块学说源于地图阅读；用制作地震分布图去寻找断裂带；战场分析和作战决策要在地图上进行，等等。再者，人的信息获取的主渠道是视觉，而这些信息中又有一半以上是与空间位置有关的，即空间信息。而地图学则从来就是研究空间信息处理与显示的科学。近十几年的 GIS 开发实践证明，凡是忽视或缺少地图学基础的各种地理信息系统，都很难在实际工作中得到应用。

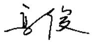
GIS 是一个地理系统的模拟系统。完全地或局部地用一个系统模拟另一个系统，从而使认识得到深化，自古以来就在科学研究中使用了。用 GIS 来模拟地理系统区别于以往的传统方法之处，

在于信息技术提供了大量数据的处理能力和建立了可更新的数据库基础。这样,就可以根据地理系统建立动力学模型,进行动态数学模拟和预测,从而减少物理模拟的盲目性,提高宏观决策与工程设计的效率和水平。在对地球空间的认识过程中,逐步以动态的、多维的概念与方法代替静态的、定性的描述来完善空间与时间的协调,这是人类空间认知能力的一大进步。在此之前,"空间"概念除了更多的是在哲学上予以争论外,在地学研究中尚无法形成对自己所研究的领域的全面的解释。那些静态分析功能,地理学家早已在地图上实现过,而动态的、多维的分析只能在空间数据库和数字制图技术支撑下的GIS中实现。

从另一个视角来看,现代GIS的最大贡献在于,在一个系统环境中既提供了启发使用者逻辑思维(建模、分析、计算等)又提供了启发形象思维(可视化、地图、图表等)的引擎,并能将二者密切结合,从而为启发使用者的创造性思维提供了极为便利的条件。

本书是华一新教授等三位青年科技专家在多年科研与教学实践的基础上撰写的,吸取了国内外大量的GIS开发的经验与方法;在内容选取上有很强的针对性;理论概念清晰;方法上强调可操作性,并针对当前一些GIS著作缺少测绘学基础的状况作了适当的补充。这是一本很具导向性的GIS专著,也是一本实用的GIS专业本科和地学领域研究生的教材。

世纪之交,祝这本总结上世纪研究成果的GIS专著在新世纪发挥它应有的作用。



2001年1月

前　言

地理信息系统(GIS)属于信息技术,可以进行空间数据的获取、处理、管理、显示、检索、分析、建模、输出等,在国民经济和国防建设的各个方面都发挥着重要的作用。本书是在作者从事本科、研究生的GIS课程教学和参加多项GIS课题研究的基础上,参阅了大量国内外有关资料而编著的。

本书共十二章,较全面地介绍了GIS的原理与技术,并顾及了GIS的最新进展,可以作为地理学和测绘学的本科和研究生的教材,并可供从事GIS研究、开发、应用的研究人员和工程技术人员参考。全书的结构为:第一章,GIS的基本概念。介绍了GIS的定义、发展、组成、应用和相关学科,以及GIS数据源的类型和特征;第二章,GIS的空间数据结构。介绍了GIS的矢量数据结构和栅格数据结构;第三章,GIS的空间数据库和数据模型。介绍了空间数据库和空间数据模型的基本概念,以及空间数据库的设计和建立;第四章,空间数据的数学基础。介绍了空间数据的数学基础,以及常用的地图投影和地图投影的变换方法;第五章,空间数据的输入。介绍了属性数据的编码、空间数据的采集和检核、空间数据的标准、空间元数据,还简要介绍了两种空间数据的交换格式;第六章,空间数据的质量。介绍了GIS数据质量的基本概念、常用方法、主要理论等,以及GIS中数据质量的评价方法;第七章,空间数据的处理。介绍了常用的空间数据处理算法,主要有:图形编辑、几何纠正、自动拓扑、矢量数据压缩、矢量栅格数据结构转换、空间数据的内插、图像数据的处理等;第八章,空间查询与空间分析。介绍了GIS中空间数据查询的基本内容,以及主要的空间分析功

能，如统计分析、地形分析、叠置分析、缓冲区分析、泰森多边形分析、网络分析、空间数据的量算等，同时还介绍了空间分析的模型和模型库及其管理系统；第九章，GIS 中的地图制图。介绍了地图符号的设计与管理、地图注记与图例配置、自动制图综合、地图设计的智能化等；第十章，GIS 的开发。介绍了项目的开发方法，以及工具型 GIS 和应用型 GIS 的开发过程；第十一章，RS、GPS 与 GIS 的结合。介绍了 RS 和 GPS 的基本概念，以及与 GIS 的结合方法和作用；第十二章，网络 GIS。介绍了计算机网络的基本概念、网络技术在 GIS 中的应用、Web GIS 的解决方案等。

编著者

目　录

第一章　GIS 的基本概念

§1.1　GIS 的定义

随着计算机技术的迅速发展和“信息高速公路”的逐步建立，人类社会已步入了信息时代，信息已成为社会赖以发展的重要基础，信息技术是高新技术之首。首先，我们对有关信息的几个概念作简要介绍。

1. 信息

信息是近代科学的一个专门术语，是客观世界中继物质和能量之后的第三个现代科学的基本概念，已广泛应用于社会的各个领域。

狭义的信息论将信息定量地定义为“两次不定性之差”，即指人们获得信息前后对事物认识的差别。例如，抛一枚硬币，这枚硬币是正是反，就有 1 比特(bit)的信息量。同理，计算机中的一个二进制位可以是 0 或 1，也具有 1 比特的信息量。

广义的信息论认为，信息是主体(人、生物或机器)与外部客体(环境、其他人、生物或机器)之间相互联系的一种形式，是主体和客体间一切有用的消息或知识，是表征事物特征的一种普通形式。

在本书中，将信息定义为——信息是客观事物的反映，它提供了客观事物的消息与知识。

2. 数据

数据是一种载存信息的物理符号。也就是说，信息可以由数据来表示，可以以数字、符号、字母等形式记载下来。

信息和数据是不可分离的，信息由数据表达，数据包含的意义就是信息。

数据是信息的载体，但数据并不就是信息，只有理解了数据的含义，对数据作出了解释，才能得到数据所包含的信息。

要从数据中得到信息必须经过处理和解释，处理是指对数据的收集、筛选、排序、转换、检索、计算、分析等，数据处理的目的是为了解释，而数据的解释需要知识和经验，不同的解释与不同的背景、目的有关。

在本书中，可以把信息和数据当作同义词。

3. 信息系统

系统是由相互作用和相互依赖的若干组成部分结合而成的，具有特定功能的有机整体。

信息系统是加工和处理信息的系统，可以对信息进行采集、处理、传输、管理、检索、分析等，可以为决策过程提供有用的信息。

本书讨论的是以计算机为主要工具的信息系统。从计算机科学的角度看，信息系统由计算机硬件、软件、数据和用户四大部分组成。信息系统具有四大功能，即采集、管理、分析、表达。

4. 空间信息

要谈到地理信息系统就必然会涉及到空间信息的概念。空间信息是指与所研究对象的空间定位与地理分布有关的信息，即指研究对象的位置、数量、质量、分布特征、相互联系与制约等。空间信息由空间数据表达。

简单地说，所谓空间信息就是指具有定位数据的信息。例如，大到世界各大洲、各国家的地理分布，小到一幢房屋，一棵树，由于它们有定位数据(如用 X、Y、Z 坐标表示)，所以都属于空间信息的范畴；而图书情报、财务等方面的信息，由于没有定位数据，就不属于空间信息。人类所使用的各种信息中，空间信息是其重要的组成部分，空间信息的分析利用在日常生活和军事行动中都已发挥了重要的作用。

空间信息包括空间位置、空间分布、空间形态、空间关系、空间相关、空间统计、空间趋势、空间对比和空间运动等方面的信息。空间位置是描述空间物体的个体位置的信息；空间分布描述的是空间对象的群体定位信息；空间形态是描述空间物体个体的形状和结构的信息；空间关系是指基于位置和形态的空间物体之间的关系；空间相关是空间物体之间基于属性数据的关系；空间统计描述的是空间物体的数、质量信息；空间趋势描述的是物体(现象)的空间分布趋势(规律)；空间对比描述空间物体在数量、质量和形态等方面对比的信息；空间运动描述空间物体随时间的变化和空间物体在空间域上的变迁和转移。

5. 空间信息的使用

在我们的生活中，有时会遇到这样一些问题：

· 一年来郑州—洛阳高速公路上的交通事故发生的位置在哪里？

· 在距伊河路小学 10 分钟步行距离内有多少二室一厅的房子出售？

· 目前，河南省哪些地方是移动电话的盲区？

回答这些问题，传统解决方案是使用一张纸质地图，在上面标注有关专题数据库中的数据，然后在地图上进行分析。为什么数据库中的数据能标注到地图上呢？这是由于存储在数据库中的信息85%以上都具有地理属性。比如，商店地址、客户地址、差转台坐标等。

如果将数据库中的属性数据叠加到电子地图上，并且将地图上的图形对象与属性数据建立联接关系。这样就可实现地图与数据库的双向查询，再增加一些地理分析功能后，就能在计算机上快速、准确地解答上面的问题了。

上述问题均可以通过对空间信息的分析和利用来解决。在现代社会中，空间信息的分析和利用已越来越广泛。下面介绍几个例子。

在高技术条件下，飞机已成为主要的作战兵器。尽管隐形飞机正在发展，但与之相对抗的雷达等防空系统也在发展，如何保证作战飞机能以最安全的路线和方式通过敌人的防空系统呢？在海湾战争中，美国空军电子安全指挥部研制的 IMOM 软件，可以在标明伊拉克防空情况的电子地图上，通过分析自动确定并显示最安全的飞行线路图，并可为飞行员指明何时、如何干扰敌方雷达，何处是敌方雷达因地形阻拦而形成的盲区，敌方武器红外或热寻的装置和监视设备的有效范围等，保证了作战飞机以最安全的路线和方式突入敌区，实施空袭。

我国许多城市建立的“110”报警台，以及与之相应的公安指挥自动化系统都涉及到空间信息的分析利用。当接到报警电话后，公安指挥自动化系统可通过电话号码的定位信息，自动显示事发地区的平面图；根据案件的发生地点、类型等，自动确定需要封锁的半径和路口；根据案发地区的警力状况（如人员、武器配置等），迅速调集警力，完成治安任务。

在经济建设中涉及到的空间信息的分析和利用的情况就更多。例如在市政建设中，在进行旧城改造，开辟或拓宽道路时，如果没有关于地下管网的准确的空间信息，就难以避免煤气管道、电缆等被挖断的情况，往往会造成重大损失。

即使在日常生活中，需要空间信息并对其加以分析利用的情况也比比皆是。当你从火车站出来，面对一个繁华而陌生的城市时，如果有一个道路咨询信息系统，那么你只要按一下电纽，就可知道你所处的位置，只要你给出需要到达的地方，它就会告诉你最便捷的交通路线和方法，等等。

以上介绍的几个例子中的共同特征是，它们都需要对空间信息根据特定的目的和要求进行分析和利用。面地理信息系统就是用子进行空间信息的分析和利用的。

6. 地理信息系统

地理信息系统，简称 GIS（Geographic Information System），

至今没有一个正式或统一的定义，以下是一些资料的定义：

地理信息系统，是在计算机软件和硬件的支持下，运用系统工程和信息科学的理论，科学管理和综合分析具有空间内涵的地理数据，以提供对规划、管理、决策和研究所需信息的技术系统（地理信息系统概论，黄杏元等，高等教育出版社，1989）。

地理信息系统，是反映人们赖以生存的现实世界的现势与变迁的各类空间数据及描述这些空间数据特征的属性，在计算机软件和硬件的支持下，以一定的格式输入、存储、检索、显示和综合分析应用的技术系统（地理信息系统原理和方法，边馥苓 编著，测绘出版社，1996）。

地理信息系统，是一种由硬件、软件、数据和用户组成的，以研究地理或地学数据的数字化或图形化采集、存储、管理、分析、表达的计算机支持系统（应用地理信息系统设计与实现，马智民等，西安地图出版社，1996）。

地理信息系统是一种采集、存储、管理、分析、显示与应用地理信息的计算机系统，是分析和处理海量地理数据的通用技术（地理信息系统导论，陈述彭等，科学出版社，2000）。

Geographic information system are computer — based systems that are used to store and manipulate geographic information（GEOGRAPHIC INFORMATION SYSTEMS: A Management Perspective, Stan Aronoff, WDL Publications, Ottawa Canada, 1991).

A Geographic System （GIS） is a computer — based information system that enables capture, modeling , manipulation, retrieval, analysis and presentation of geographically referenced data (GIS—A Computing Perspective, Michael F. Worboys, Taylor&Francis, 1995).

英国著名地理信息系统和地图学家 D. Rhind 说："如果问 100 个人，什么是地理信息系统，也许会有 99 种答案，但是地理信息系

统是一种技术。"从学术上讲，对 GIS 有三种观点：

(1)地图观——主要来自景观学派和制图学派，认为 GIS 是地图显示、处理与使用的系统，每个数据集被看作是地图、图层、要素等；

(2)数据库观——主要来自于计算机学派，强调优化设计、数据库技术和有效存取数据的重要性；

(3)空间分析观——主要来自地理学派，强调空间分析与模拟的重要性，并提出了地理信息科学的概念。

对于一个迅速发展的、尚未完全成熟的领域过早地追求一个完善的定义，并非是明智之举。只要把握其基本特征和基本构成即可。因此，我们真正需要掌握的是 GIS 有什么用途？GIS 的结构是怎样的？有那些关键技术？如何开发和应用？等等。

简单地说，地理信息系统就是一门处理地理空间数据的技术。

"数据"是信息的具体表示方式。信息用数据描述后，计算机才能存贮和处理。

"地理"意味着数据是参照于地球的。

"空间"意味着数据所表示的事物是具有一定的点位、形状、性质等特征的。

"处理"是指用计算机对数据进行输入、管理、查询、分析等操作。

所以，地理信息系统就是对地理空间数据进行输入、管理、查询、应用的技术，因此其应用领域极为广阔，以上所举的几个例子只是地理信息系统在某些方面的应用。

§1.2 GIS 的发展

地理信息系统是处理地理空间信息的高技术，而地理空间信息并不是什么新东西，早在几千年前就表示在古地图上了，到目前为止，地图仍然是表示地理空间信息的最佳媒体。

最早的计算机是用来处理数字的，主要目的是解决数学问题，然后逐步发展为可以表示字符和文字了，利用数字和字符就可表示地理空间信息。但一开始（自20世纪60年代起）用计算机处理地理空间信息的目的是为了以计算机为手段进行地图生产，即开发计算机地图制图系统。在计算机地图制图系统的基础上，加拿大研制成功了第一个地理信息系统（CGIS，1971）用于自然资源的规划和管理。

由于计算机硬软件技术的飞速发展，地理信息系统朝着实用方向迅速发展起来。一些发达国家建立了许多专业性的地理信息系统，我国从20世纪80年代开始也着手GIS方面的研究，但与发达国家相比，差距仍然很大。

1. 国际上GIS的发展概况

（1）20世纪60年代开始发展。随着社会的进步，对于自然资源和环境的规划、管理等的要求越来越迫切，随着计算机技术的发展，促使了对地图的分析、应用和输出系统的研制。

第一个综合性的地图制图软件是美国哈佛大学 Howard T. Fisher 等研制的 SYMAP（Synagraphic Mapping System，1963）。

随着有关对空间数据的采集、存贮、检索、分析、显示和输出技术的发展，最终导致了GIS的产生。

1963年，加拿大着手建立加拿大地理信息系统（CGIS），用于处理和应用加拿大土地调查所获得的大量数据，该系统于1971年正式投入运行，是国际上公认的、较为完善的、大型GIS。

同时，有关的组织和机构纷纷建立。

· 1966，成立美国城市和区域信息系统协会（URISA）。

· 1969，成立美国州信息系统全国协会（NASIS）。

· 1968，国际地理联合会（IGU）成立了地理数据采集与处理委员会（CGDSP）。

这些组织相继举行了一系列GIS的国际研讨会，对GIS技术的发展和传播起到了重要的指导作用。

这一时期的 GIS 软件主要以计算机制图为主，解决了计算机制图方面的有关问题，如：

· 栅格数据与矢量数据的互相转换

· 自动拓扑

· 地图的拼幅

· 多边形的合并，叠加

· 按属性查询空间数据

· 地图和表格的输出

(2)20 世纪 70 年代，巩固时期。世界各地研制了不同专题、不同规模、不同类型的许多 GIS，美、加、英、德、瑞典、日本等国投入了大量的人力、物力、财力。如：

· 美国林业调查局，开发了资源信息显示系统 RIDS

· 日本建立了国土信息系统

· 瑞典在中央、区域和市三级建立了许多 GIS，如城市规划 GIS

· 法国建立了地理数据库系统 GITAN

以遥感为数据源的 GIS 逐渐受到重视。

1976，美国喷气推动实验室(JPL)研制成功兼有影像数据处理和 GIS 功能的影像信息系统，IBIS(Image Based Information System)。

1979～1980，NASA 的地球资源实验室开发了 ELAS 地理信息系统，可处理卫星影像数据和地图数据，并用于生产专题图。

70 年代，大约有 300 多个 GIS 软件投入使用。

同时成立了一些组织机构，进行 GIS 方面的研究。

1975 年，国际地理联合会，成立了地理数据传输与处理委员会，调查与空间数据处理有关的计算机软件。

1978 年，国际测量工作者联合会(FIG)，决定它的第三委员会的任务是研究 GIS。

1980 年，美国地质调查局出版了美国 GIS 专家 D. F. Marble

等的《空间数据处理计算机软件》的报告，总结了1979以前世界各国所能收集到的GIS及有关的计算机软件的研制情况。

这一时期的GIS软件在技术方面的发展特点为：

·属性数据和空间数据分别管理

·地图扫描输入技术有一定进展

·人机交互技术成熟

·研制了大量的数据处理和GIS软件

(3)20世纪80年代，突破阶段。随着图形工作站和微机性能价格比的迅速提高，计算机技术的应用更加广泛。GIS软件从简单的单一功能，分散的系统，发展到多功能、数据共享的GIS，并向智能化发展。

·GIS开始全面应用

·已推广到第三世界国家

·与RS的结合，开始用于全球性问题，如全球沙漠化、核扩散、核废料等。

·出现了大量的商业性GIS工具软件，形成GIS产业。1989年，市场上有报价的GIS工具软件就有70多个。

·GIS由机助制图性质向分析应用转变，用于更复杂的领域，如城市化的研究等。

·GIS成为许多国家和部门的必备工具。

在测绘界一些GIS刊物创刊，一些杂志改名。

1987，英国创刊《国际地理信息系统杂志》

1988，德国创刊《土地信息系统》

1980第一期：美国的《测量与制图》更名为《测量与土地信息系统》

《美国地图制图员》更名为《地图制图与土地信息系统》

这一时期GIS软件在技术上的特点为：

·栅格数据矢量化的研究提高了数据输入的效率

・可支持不同形式的地图输出

・可处理更复杂和数据量更大的情况

・空间分析和应用能力大大提高

・空间数据库与属性数据库的无缝结合

(4)20 世纪 90 年代,全面应用。随着地理信息产业的建立和数字化信息产品在全世界的普及,GIS 必将深入到各行各业、千家万户,成为人们不可缺少的工具和助手,并以此来迎接信息时代的到来。

在这一阶段,GIS 发展的特点如:

・集成化和智能化

・GIS、GPS、RS 的结合

・GIS 理论与技术的进一步完善(数据结构、模型、质量…)

・标准化和数据共享

・GIS 的网络化

・GIS 的更新换代(新的操作系统、语言、工具、设计思想…)

2. 我国 GIS 的发展概况

(1)20 世纪 70 年代,准备阶段。从 20 世纪 70 年代初,我国开始探讨计算机在地图制图和遥感领域中的应用。

1976 年,召开了第一次遥感会议。

1977 年,诞生了我国第一张由计算机输出的全要素地图。

1978 年,召开了第一次全国数据库学术讨论会。

(2)20 世纪 80 年代,起步阶段。在 80 年代,我国的 GIS 研究与开发由自发的研究与试验,开始进入有组织、有计划的发展阶段。

开始着手 GIS 规范的研究,建立了一些专题的 GIS,并开始了 GIS 方面的学术讨论。

(3)20 世纪 90 年代,稳步发展。90 年代以来,我国建立了相当数量的 GIS 运行系统,在国民经济建设和国防建设中发挥了重要作用。一批高等院校开设了与 GIS 相关的新专业,培养了一大批

从事 GIS 研究与开发的高层次人才。具有我国自主版权的 GIS 基础软件的研制逐步进入了产业化轨道，部分软件已提供用户使用。1994 年中国 GIS 协会成立，并举办了一系列的 GIS 讨论会。一批与 GIS 相关的标准和规范已经公布执行。所有这些都标志着我国的 GIS 正在以稳固的步伐前进着。

§1.3 GIS 的组成

GIS 的组成可由图 1—1 所示。

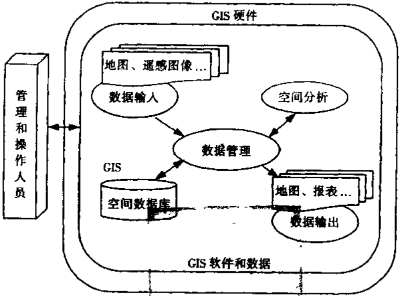


图 1--1 GIS 的组成

GIS 由四部分组成，即计算机硬件系统、计算机软件系统、地理数据和系统的管理与操作人员。

1. 计算机硬件系统

GIS 的硬件配置通常包括四个部分：

(1)计算机主机:包括屏幕、键盘、鼠标等。

(2)输入设备:包括数字化仪、扫描仪、测绘仪器等。

(3)存贮设备:包括软盘、硬盘、磁带和光盘等。

(4)输出设备:包括显示器、绘图机、打印机等。

2. 计算机软件系统

GIS 的计算机软件系统可以按功能分为五个部分:

(1)数据输入。数据是GIS 的血液。数据输入的目的是将现有的地图、外业观测成果、航空像片、遥感图像、文本资料等转换成GIS 可以处理与接收的数字形式,使系统能够识别、管理和分析。通常要经过验证、修改、编辑等处理。不同数据输入需要用到不同的设备。

例如,对于文本数据通常用交互的方式通过键盘录入,也可用扫描仪扫描后用字符识别软件自动录入;对于矢量地图数据,可用平板数字化仪,采用手扶跟踪的方法输入,也可用扫描仪扫描成图像后,用栅格数据矢量化的方法自动追踪输入;等等。

GIS 软件的这一部分还应具有数据转换装载的功能,即能把其他 GIS 或专题数据库中的数据通过转换装载到当前的 GIS 系统中。

这一部分 GIS 软件的数据处理工作主要是几何纠正、图形和文本数据的编辑、图幅的拼接、拓扑关系的生成等,即完成 GIS 的空间数据在装入 GIS 的地理数据库前的各种工作。

(2)数据管理。数据管理部分是GIS 的心脏。在GIS 的核心部分,有一个巨大的地理数据库,它必须能管理存储于 GIS 中的一切数据,具有数据库的定义、维护、查询、通讯等功能。

GIS 的地理数据库中主要存贮有空间数据、专题数据,以及多媒体数据等。目前,GIS 的地理数据库主要用商用关系数据库进行管理,或者在关系数据库之上建立一面向对象层,以建立面向对象的 GIS。

GIS 的地理数据库管理系统应提供数据库的定义、管理、维

护、通讯等功能。所以，GIS 的地理数据库不仅要能存贮和管理空间数据、专题数据、拓扑关系数据等，还应能进行数据压缩、数据转换、空间查询等，如位置查询、属性查询、拓扑查询等等。

(3)空间分析。空间分析是 GIS 的大脑，它是 GIS 区别于一般事务数据库和其他一些系统的重要特征。它通过对 GIS 中空间数据的分析和运算，为 GIS 的具体应用提供分析处理后的信息。但 GIS 基础软件(或称工具型 GIS)往往只提供最通用、最基本的分析模型，各种专业的应用模型需要专业人员进行二次开发。GIS 通常提供地形分析、叠置分析、缓冲区分析、网络分析等基本的分析功能。

(4)数据输出。将 GIS 中的数据经过分析、转换、处理、组织，以某种用户可以理解的形式(如报表、地图等)提供给用户，网络 GIS 还可以直接通过 Internet 传给网络用户。这就是 GIS 数据输出部分的任务。

由于地图是地理信息表示的最佳媒介，所以，GIS 的数据输出中最大量的是地图的显示与输出。输出的产品可以是在屏幕、打印机、绘图机、磁盘等上面进行显示和保存，也可以通过网络直接传递给其他用户。

(5)应用模块。这是指应用于某种特定任务的 GIS 软件模块，在 GIS 的基础软件(或称工具软件)中是不具备的，但往往是用户最关心的。如房地产管理、自然灾害分析等。

为了有效地应用 GIS，必须在 GIS 基础软件的基础上，开发出相应的应用模块。这部分软件往往由用户自行开发或委托专业人员开发，通常需要用到数学建模和专家系统等技术。

这部分软件可以根据不同的用户目的，导致完全不同的功能，如城市规划的应用模块和水资源调查的应用模块，在功能上就完全不一样。GIS 的这一部分是为用户的事务性工作直接服务的，往往带有辅助决策的性质。

GIS 软件各部分的关系如下图(图 1—2)所示。

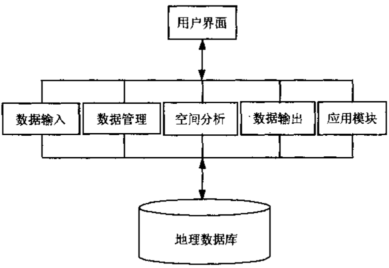


图 1—2 GIS 软件系统的组成

3. 地理数据

数据是 GIS 的血液，没有数据 GIS 就无法运行。地理数据是具有空间定位的自然、社会、人文、经济等方面的数据，可以是图形、图像、文本、表格、数字、声音等，是用户通过数字化仪、扫描仪等设备输入到 GIS 的地理数据库中的。

GIS 中数据的质量（精度、现势性等）直接影响到 GIS 分析结果的可靠性。由于数据是随着时间不断老化的，为了使 GIS 能有效、可靠地发挥作用，GIS 地理数据库中的数据必须被不断地更新。更新数据就需要投入，因此，GIS 是需要不断的资金投入的。否则，如果 GIS 中的数据是"垃圾"，那么其输出的结果也将是"垃圾"，不会因为昂贵的设备和高级的技术人才而改变数据的质量。

GIS 的数据输入部分主要把存在于地图、遥感像片等载体的空间数据，把存在于专题数据库或以各种文字、图像、声音和统计数据等形式存在的属性数据输入到系统中，使系统能够识别、管理和分析。

GIS 中有两种数据，一种是空间数据，另外一种是属性数据。空间数据是描述目标的空间位置、几何形态以及与其他目标的空间关系的数据。例如描述一幢房子位置和形状的坐标数据。空间数据通常是通过测绘手段获取的（包括地图、遥感、GPS 等）。属性数据是指描述空间目标的社会或自然属性的数据，如房子的户主、建筑年代、建筑材料等。在建立 GIS 的最初阶段，数据采集和转换的经费往往高达整个系统的 70%～80%。

4. 管理和操作人员

任何先进技术的引进和开发应用，都必须拥有掌握该技术的人才。对 GIS 的管理和操作人员，应对他们进行培训，使他们能以与传统手工操作不同的、新的方式进行思维和工作。

从事 GIS 的工作人员通常可分为以下几类：

• 低级技术人员——不必知道 GIS 如何工作，任务是数据的输入、结果的输出等。

• 业务操作人员——应熟练掌握 GIS 的操作，维持 GIS 的日常运行，完成应用任务。

• 软件技术人员——必须精通 GIS，负责系统的维护、系统的开发、数学模型的建立等。

• 科研人员——利用 GIS 进行科研工作，并能提出新的应用项目、新的要求和功能等。

• 管理人员——包括决策、公关等人员，应懂得 GIS 技术，能介绍 GIS 的功能，寻找用户等。

如果只是盲目地引人 GIS 软件，买一些硬件设备，由几个热心的人进行开发、管理和使用，是难以产生应有的效益的，必须有稳定的组织机构来进行 GIS 的开发、管理和使用。

§1.4 GIS 的类型

GIS 可以根据不同的分类指标，形成多种分类体系。GIS 常见

的分类如下。

1. **工具型** GIS **与应用型** GIS

根据 GIS 的功能,GIS 可分为工具型和应用型两种。

工具型 GIS——常称为 GIS 工具、GIS 开发平台、GIS 外壳、GIS 基础软件等,没有具体的应用目标,通常为一组具有 GIS 功能的软件包,是建立应用型 GIS 的支撑软件。如 ARC/INFO、MapInfo 等。

应用型 GIS——具有具体的应用目标,一定的规模,特定的服务对象,特定的数据和用户。通常,应用型 GIS 是在工具型 GIS 的支持下建立起来的。

工具型 GIS 是 GIS 研究和开发的核心内容。工具型 GIS 是一组具有图形图像数字化、数据管理、查询检索、分析运算和制图输出等 GIS 基本功能的软件包,通常能适应不同的硬件条件,软件的功能强、性能稳定。只要在工具型 GIS 中加入地理空间数据,开发有关的应用模型和界面,就可成为一个应用型的 GIS 了。

在通用的 GIS 工具(GIS 基础软件)支持下建立应用型 GIS,可节省大量的软件开发费用,缩短系统的建立周期,提高系统的技术水平,使开发人员能把精力集中于应用模型的开发,且有利于标准化的实行。

2. **专题 GIS 与综合 GIS**

根据研究对象的性质和内容,可以把应用型 GIS 分为专题 GIS 和综合 GIS。

专题 GIS——具有有限的目标和明确的专业特点,是为特定的专门目的服务的。如水土流失信息系统、草场资源信息系统等。

综合 GIS——以区域综合研究和全面的信息服务为目标,可以有不同的规模,如国家级、省级、县级等。综合 GIS 的例子如加拿大国家信息系统等。

3. **矢量型 GIS、栅格型 GIS 和混合型 GIS**

根据 GIS 的数据结构,可以把 GIS 分为矢量型 GIS、栅格型

GIS 和混合型 GIS。

矢量型 GIS——当空间数据是由矢量数据结构表示，即用坐标值来精确地表示点、线、面等地理实体时，这种 GIS 称为矢量型 GIS。

栅格型 GIS——当空间数据是由栅格数据结构表示，即以规则的像元阵列来表示空间地物或现象的分布时，这种 GIS 称为栅格型 GIS。

混合型 GIS——指矢量、栅格数据结构并存的 GIS。

§1.5 GIS 的应用

1. GIS 的主要作用

大多数功能齐全的 GIS 具备三种类型的作用。

①制图——对于 90%的用户，最重要的应用是制图，制作各种专题地图。

②查询——这是 GIS 的第二大应用。用户可以将地图作为查询空间数据库的手段。如果没有 GIS 的帮助，查询的对象只能是表格或清单，如果有 GIS，用户可以在地图上进行查询，看到表示于地图上的查询结果。

③分析——GIS 的第三大应用在于空间分析。空间分析是指为制定规划和决策，应用逻辑或数学模型分析空间数据。

2. GIS 的应用

用地图表示空间数据的方法至少已有 2000 年了，但只有通过 GIS，空间数据才真正全面地发挥了它的各种作用。GIS 既可以对数据进行定量分析，也可以把分析的结果表达为图形，同时支持数据思维和空间思维，比起在传统地图上的空间分析和针对统计数据的定量分析，已有了质的改进，因而在科学研究、管理决策等各方面都有广泛的应用。有关 GIS 的应用，举例说明如下：

(1)对全球动态监测。在全球范围内，GIS 可用于全球变化的

动态监测。如联合国粮农组织(FAO)在意大利建立了遥感与GIS中心，负责对欧洲和非洲的农作物生产的病虫害防治提供实时的监测技术服务。

(2)为国家基础产业的管理和决策服务。在一个国家(或一个大区域)的范围内，可用于全国范围的自然资源调查、环境研究、土地利用调查、森林管理、农作物生产、各种灾害预测和防治、国民经济调查和宏观决策分析等。

(3)为城市建设服务。在一个城市范围内，GIS可用于土地管理、房地产经营、污染治理、环境保护、交通规划、管线管理、市政工程服务和城市规划等。

(4)为企业的生产经营和工程建设服务。在一个企业范围内，可用于生产经营和管理。如德国的露天煤矿，利用航测方法建立和更新矿区的GIS，在此基础上设计开采面作业计划、矿石运输线路、废矿石堆放位置等，从而使生产达到最佳作业化。

在香港新机场的建设中，通过采用数字测量系统(DPS)和GIS，飞机每天摄影一次，GIS在24小时内给出土方量，并计算出工时等各种费用，以及机场的基本情况，用以指挥机场的施工。

(5)在军事上的应用。一切战略的、战役的或战术的行动都离不开战场的地理环境，诸军(兵)种联合作战更是如此，而GIS在获取、存储、处理和分析空间地理环境信息并辅助决策方面具有特殊的地位和作用，西方国家对这一领域的研究和应用表现出了极大的兴趣，并给予了高度的重视。

在海湾战争这场高技术、高强度的现代局部冲突中，美军积极采用GIS技术，及时为"沙漠盾牌"和"沙漠风暴"行动提供保障，主要用于模拟部队和车辆的机动性，估算化学武器影响的危险区域，模拟用烟雾遮蔽战场的效果，评定地形对武器性能的影响，通过对策模拟制定完善的军事行动计划，提供水源探测所需点位等。

美国国防制图局(DMA)的战场GIS，是在工作站上建立的，它能用自动影像匹配和自动目标识别技术，处理卫星和高低空侦

察机实时获得的战场数学影像，及时地（不超过 4 小时）将反映战场现状的正射影像图叠加到数字地图上，数据直接传送到前线指挥部和五角大楼，为军事决策提供 24 小时的实时服务。

（6）日本政府利用 GIS 技术建立防震减灾系统。日本横滨大地震后，日本建筑署建设研究所等政府机构在联合国区域发展中心的支持下，采用 ARC/INFO 软件对横滨大地震的震后影响作出评估，建立了各类数字地图库，包括地质断层、倒塌建筑等，把各图层进行叠加分析，得出对应急有价值的信息。该系统的建成可以对象神户这样的大城市在大地震后作出快速响应，最大限度的减少伤亡和损失。

（7）ERDAS 公司的遥感和图像 GIS 软件 Imagine 协助波黑和平谈判。1996 年初，在美国俄亥俄州的代顿穆塞克三方代表举行了举世瞩目的波黑和平谈判，三方土地边界的确定是谈判的关键所在。为了加速谈判进程，美国国防制图局采用了 Imagine 软件。Imagine 是一种图像分析和地形可视化 GIS 工具，它能使每位谈判者能快速直观和准确地看到每次边界的变化是如何分割道路和城市以及影响自然资源的分配的。谈判代表在酸酯地图上画出他们建议的边界，国防制图局的技术人员只需几分钟就可在 Imagine 系统的数字地图上输入新的边界，并输出新的地图供谈判使用。在三周的时间里，共修改和输出 3000 多幅地图，最终的和平协议里就包含了 100 多幅有三方总统签字的地图。

（8）欧洲政府机构利用 GIS 管理全欧环境计划。1998 年 3 月，由欧共体支持的企业协会开始利用 Internet 网络和基于 Internet 的 GIS 软件——Autodesk 公司的 MapGuide，向商家和政府机构提供关键信息以解决环境问题。欧洲的环境问题多种多样，法兰西（保护 Lyon 地区空气和水的质量），意大利（Arno 河盆地洪水泛滥），希腊（旅游对环境的影响），苏格兰（土地使用和农业问题），使用网络 GIS 技术，全欧洲的地区政府和机构可以共享地图和基于地图的信息。

(9)世界动物保护基金会利用 ARC/INFO 保护野生动物。世界动物保护基金会利用 ARC/INFO 作为它“老虎与人类”保护项目的主要技术工具。它采用 ARC/INFO 在实地进行数据采集，分析老虎的重要栖息地，对老虎的种类、数量、猎物、生态系统等进行综合分析，并把分析的结果和地图发布给当地政府和其他组织，为他们制定和实施法律提供帮助。

GIS 的应用领域极为广泛，如可用于地图制作与生产、区域地质调查、矿产资源调查、基础地质研究、环境评价、城市规划、灾害监测、公安侦破、军事指挥、土地调查、地籍管理、市政设施管理等与空间信息有关的众多领域，以上只是一个概略的介绍。

§1.6 GIS 的数据源

GIS 的数据源，是指建立 GIS 的地理数据库所需的各种数据的来源，主要包括地图、遥感图像、文本资料、统计资料、实测数据、多媒体数据、已有系统的数据等。

1. 地图数据

地图是 GIS 的主要数据源，因为地图包含着丰富的内容，不仅含有实体的类别和属性，而且含有实体间的空间关系。地图数据主要通过对地图的跟踪数字化和扫描数字化获取。地图数据不仅可以作宏观的分析(用小比例尺地图数据)，而且可以作微观的分析(用大比例尺地图数据)。在使用地图数据时，应考虑到地图投影所引起的变形，在需要时进行投影变换，或转换成地理坐标。

地图数据通常用点、线、面及注记来表示地理实体及实体间的关系，如：

点——居民点、采样点、高程点、控制点等。

线——河流、道路、构造线等。

面——湖泊、海洋、植被等。

注记——地名注记、高程注记等。

2. 遥感数据

遥感数据是GIS的重要数据源。遥感数据含有丰富的资源与环境信息，在GIS支持下，可以与地质、地球物理、地球化学、地球生物、军事应用等方面的信息进行信息复合和综合分析。遥感数据是一种大面积的、动态的、近实时的数据源，是GIS数据更新的重要手段。

3. 文本资料

文本资料是指各行业、各部门的有关法律文档、行业规范、技术标准、条文条例等，如边界条约等。这些也属于GIS的数据。

4. 统计资料

国家和军队的许多部门和机构都拥有不同领域(如人口、基础设施建设、兵要地志等)的大量统计资料，这些都是GIS的数据源，尤其是GIS属性数据的重要来源。

5. 实测数据

野外试验、实地测量等获取的数据可以通过转换直接进入GIS的地理数据库，以便于进行实时的分析和进一步的应用。GPS(全球定位系统)所获取的数据也是GIS的重要数据源。

6. 多媒体数据

多媒体数据(包括声音、录像等)通常可通过通讯口传入到GIS的地理数据库中，目前其主要功能是辅助GIS的分析和查询。

7. 已有系统的数据

GIS还可以从其他已建成的信息系统和数据库中获取相应的数据。由于规范化、标准化的推广，不同系统间的数据共享和可交换性越来越强。这样就拓展了数据的可用性，增加了数据的潜在价值。

§1.7 GIS的数据特征和数据类型

1. 空间数据的特征

空间数据具有三个基本特征：

(1)属性特征——用以描述事物或现象的特性，即用来说明"是什么"，如事物或现象的类别、等级、数量、名称等。

(2)空间特征——用以描述事物或现象的地理位置，又称几何特征、定位特征，如界桩的经纬度等。

(3)时间特征——用以描述事物或现象随时间的变化，例如人口数的逐年变化。

由于空间数据具有上述特征，所以在 GIS 中的表示是非常复杂的。目前的 GIS 还较少考虑到空间数据的时间特征，只考虑其属性特征与空间特征的结合。实际上，由于空间数据具有时间维，过时的信息虽不具有现势性，但却可以作为历史性数据保存起来，因而就会大大增加 GIS 表示和处理数据的难度。

2. 空间数据的类型

根据空间数据的特征，可以把空间数据归纳为三类：

(1)属性数据——描述空间数据的属性特征的数据，也称非几何数据。即说明"是什么"，如类型、等级、名称、状态等。

(2)几何数据——描述空间数据的空间特征的数据，也称位置数据、定位数据。即说明"在哪里"，如用 X、Y 坐标来表示。

(3)关系数据——描述空间数据之间的空间关系的数据，如空间数据的相邻、包含等，主要是指拓扑关系。拓扑关系是一种对空间关系进行明确定义的数学方法。

§1.8 GIS 的相关学科和技术

1. GIS 涉及的学科和技术

GIS 是传统科学与现代技术的结合，涉及到多门学科和技术。图 1—3 列出的是一些与 GIS 关系密切的学科与技术。

地理学广泛涉及到人类居住的地球与世界。地理学中的空间分析历史悠久，是 GIS 中空间分析方法的基本来源。

地图学有着古老的历史，研究空间信息的表示方法。地图是空间信息最有效的表示方法，在GIS中也如此。地图不仅是GIS的重要数据源，而且也是GIS的主要输出方式。GIS是在计算机地图制图的基础上发展起来的，是否具有空间分析功能是区分GIS与计算机地图制图的主要标志。

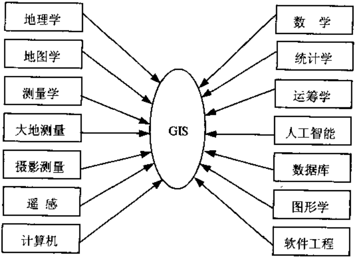


图1—3 GIS涉及的学科的技术

测量学可以为GIS提供高质量的、准确的位置信息，如土地边界、建筑物的位置等。

大地测量是GIS中精确定位信息的来源。GPS(全球定位系统)技术可以方便、实时地获取三维的空间数据。

摄影测量是GIS中空间数据的又一重要来源，尤其是地形数据。

遥感(RS)是GIS中更新空间数据的有效和主要的方法，也是GIS的重要数据源，同时，GIS又可进一步支持遥感信息的开发和应用。

计算机技术是建立GIS的基础，因为GIS是一种基于计算机

的信息系统。

人工智能技术为GIS作为决策支持的工具提供了理论、技术和方法。

数据库技术为海量空间数据的管理提供了有效的技术手段，地理数据库是GIS的核心。

计算机图形学为GIS中空间数据的操作和运算提供了直接的方法。

软件工程为GIS的开发提供了软件设计、管理等方面的方法和工具。

数学的许多分支，如图论、几何学等，已广泛应用于GIS中空间数据的分析和运算。

统计学不仅在GIS空间建模方面有重要的应用，而且也是分析GIS数据质量的基础。

运筹学为基于GIS的决策、优化提供了方法。

与GIS相关的学科和技术有很多，以上只是对关系较为密切的学科和技术作了简单介绍。

2. 关于GIS与地图学

(1)GIS拓宽了地图学的领域。德国古典哲学家(Immanual Kaut 1724～1804)曾将知识划分成三个不同的领域：

研究特定对象、对象集和现象的科学，如：生物学、几何学等；

研究按时间顺序而变化的事物的科学，如：历史学；

研究在各自空间下特征的科学，如：地理学。

Kaut的这种划分受时代和当时的科学技术所限。近年来，由于计算机技术的发展，由于学科的交叉，必然会产生新的边缘学科，地图学正处于50年来最活跃的时期，也是最不稳定的时期。在这样的时期，一些地图学家把触角伸入到一些新技术、新领域，使地图学产生了生命力旺盛的新的生长点。GIS就是这样的生长点之一。

GIS不但能够制图，而且其技术含量和功能已经远远超出地

图的范围。但这不等于说GIS完全能够取代地图,我们看到,GIS所制作出来的地图,无论是设计、制作、还是符号系统,都基本上遵循地图学的基本理论和技术。从GIS中的制图功能来看,GIS技术的价值不在于代替地图,而是空前扩大了地图的领域,增加了地图内容的深度和地图应用的宽度。

(2)GIS与地图学一脉相承。地理信息系统脱胎于60年代的机助制图系统,GIS的核心功能,即分析功能是在数字地图的基础上建立起来的,因为任何地理分析方法都离不开空间关系的参与。正是从这点出发,GIS作为地图使用的一个成功范例而成为地图学新开拓的领域之一。高俊院士在中国测绘学会第五次代表大会上列举了地图学专家进入GIS领域的优势,说明了GIS与地图学一脉相承的特点。虽然这不能理解成不准其他学科和行业涉足GIS的领域,就像地图学家从来也不反对其他领域的专家制造地图一样。特别是GIS的发展,其功能和技术环节已经超越了地图学的范围。GIS的空间信息可视化的功能虽然是地图学家提供的,但分析研究的主题和规律的探索却是由GIS用户来决定的,例如气象学家用GIS研究气象预报,土壤学家用GIS研究土壤与气候的关系等等。这就和地图学与专题地图的关系一样。把专题地图作为地图学的一部分是公认的,但大多数的专题地图作者却并非地图专家,而是各行各业的专家。地图学提供的只是最佳的表示法的理论和技术,地图专家参与GIS的开发,具有很大的优势,但任何人掌握了地图学的知识,也一样能有效的开发GIS。

(3)GIS是地图学在信息时代的发展。国际和国内的专家们就GIS的学科特点也持续争论了很长时间。但GIS是包括地图学在内的多学科的产物却得到基本的认同,另外一种倾向性的见解是地图学与GIS都是空间信息处理技术,一般认为地图学强调图形信息传输,而GIS注重空间数据处理和分析。如果把现代地图学包括了地图使用这一特点考虑进去,两者就更加密不可分。可以说GIS是地图学在信息时代的发展。

3. 地理信息系统与其他相关系统的区别和联系

(1)GIS 与一般事务数据库。举个很简单的例子:电话查号台可看作是一个事务数据库,它只能回答用户所询问的电话号码,而相关的 GIS 除了查号之外,还可以提供电话用户的地理分布,空间密度,最近的邮电局等信息。GIS 与一般事务数据库的区别在于能处理空间数据,它不仅要管理一般的数字、文字数据库,而且还要管理具有空间拓扑结构的图形数据库。

(2)GIS 与数字地图制图。数字地图制图是 GIS 的重要组成部分。数字地图制图主要是指地图在计算机中的输入、存储、管理、显示以及输出,它所强调的是地图的图形表示,通常只有图形数据而没有非图形数据,它也能进行空间查询检索,但它不能综合二者的数据进行深层次的空间分析,以提供对规划、管理和决策有用的信息。

(3)GIS 与 CAD。计算机辅助设计(CAD)系统主要提供建筑、汽车等计算机图形设计的交互手段。它在很多地方都与 GIS 很相似。当前很多原先从事 CAD 的公司都开发出了自己的 GIS,比较著名的如生产 AutoCAD 的 Autodesk 公司开发出了 MapGuide。GIS 和 CAD 两者有大量的技术重叠,两者都用计算机处理图形,都能生成漂亮的图像,具有相似的输入输出设备,在 GIS 中,大多可以实现与 CAD 的数据交换。但是,两者的相似性到此为止。下面是两者的对比:

CAD 的几何形状主要是由制图员构造,而 GIS 的几何形状是测量、数字化或扫描进去的。

CAD 中采用的多是几何坐标系,而 GIS 中采用的多是大地坐标。

CAD 中属性库的功能较弱,而 GIS 属性库的内容结构复杂,图形与属性的交互十分频繁,功能强大。

在 CAD 中,栅格数据很少用,但在 GIS 中,这是地图数据或卫星数据的一种非常有效的表示方法。

CAD的几何形状一般都是规则的图形及其组合，特征点较少，而GIS的几何形状多为不规则的自然目标，一个多边形、一条链可能有上千个顶点。因此，图形处理的难度相对较大。

第二章　GIS 的空间数据结构

空间数据结构——是指适合于计算机存贮、管理、处理的空间数据的逻辑结构。换句话说，是指空间数据以什么形式在计算机中存贮和处理。

空间数据结构分为矢量数据结构和栅格数据结构两种。

§1.1　矢量数据结构

1. 基本概念

矢量数据结构——通过坐标值来精确地表示点、线、面等地理实体的方法。

点——由一对 x，y 坐标表示。

线——由一串有序的 x，y 坐标对表示。

面——由一串有序的、且首尾坐标相同的 x，y 坐标对表示。

矢量数据结构可以表示现实世界中各种复杂的实体，当问题可描述成线和边界时，特别有效。矢量数据冗余度低，结构紧凑，并具有空间实体的拓扑信息，便于深层次分析。矢量数据的输出质量好、精度高。

矢量数据的获取方式通常有：

(1)由外业测量获得。可利用测量仪器自动记录测量成果(常称为电子手薄)，然后转到地理数据库中。

(2)由摄影测量获得。可利用摄影测量技术获取矢量化的测量成果，然后转到地理数据库中。

(3)由栅格数据转换获得。利用栅格数据矢量化技术，把栅格

数据转换为矢量数据。

(4)跟踪数字化。用跟踪数字化的方法，把地图变成离散的矢量数据。

2. 矢量数据结构的表示

矢量数据的表示方法多种多样，但基本上类似，可触类旁通。

在 GIS 中，矢量数据表示时应考虑以下问题：

- 矢量数据自身的存贮和处理。
- 与属性数据的联系。
- 矢量数据之间的空间关系(拓扑关系)。

下面分别介绍矢量数据的简单数据结构和拓扑数据结构。

(1)简单数据结构。矢量数据的简单数据结构没有拓扑关系，主要用于矢量数据的显示、输出，以及一般的查询和检索。可分别按点、线、面三种基本形式来描述。

——点的矢量数据结构

点的矢量数据结构可表示为：

标识码	X,Y 坐标

标识码：按一定的原则编码，简单情况下可顺序编号。标识码具有唯一性，是联系矢量数据和与其对应的属性数据的关键字。属性数据单独存放在数据库中。

在点的矢量数据结构中也可包含属性码，这时其数据结构为：

标识码	属性码	X,Y 坐标

属性码：通常把与实体有关的基本属性(如等级、类型、大小等)作为属性码。属性码可以有一个和多个。

X,Y 坐标：是点实体的定位点，如果是有向点，则可以有两个坐标对。

——线(链)的矢量数据结构

线(链)的矢量数据结构可表示为：

标识码	坐标对数 n	X,Y 坐标串

标识码的含义与点的矢量数据结构相同。同样,在线的矢量数据结构中也可含有属性码,如表示线的类型、等级、是否要加密、光滑等等。

坐标对数 n:构成该线(链)的坐标对的个数。

X,Y 坐标串:这是构成线(链)的矢量坐标,共有 n 对。也可把所有线(链)的 X,Y 坐标串单独存放,这时只要给出指向该链坐标串的首地址指针即可。

——面(多边形)

面的矢量数据结构可以象线的数据结构一样表示,只是坐标串的首尾坐标相同。这里介绍链索引编码的面(多边形)的矢量数据结构,可表示为:

标识码	链数 n	链标识码集

标识码的含义同点和线的矢量数据结构,在面的矢量数据结构中也可含有属性码。

链数 n:指构成该面(多边形)的链的数目。

链标识码集:指所有构成该面(多边形)的链的标识码的集合,共有 n 个。

这样,一个面(多边形)就可由多条链构成,每条链的坐标可由线(链)的矢量数据结构获取。这种方法可保证多边形公共边的惟一性;但多边形的分解和合并不易进行;邻域处理比较复杂,需追踪出公共边;在处理“洞”或“岛”之类的多边形嵌套问题时较麻烦,需计算多边形的包含等。

(2)拓扑数据结构。拓扑关系是一种对空间结构关系进行明确定义的数学方法。具有拓扑关系的矢量数据结构就是拓扑数据结构,拓扑数据结构是 GIS 的分析和应用功能所必需的。拓扑数据

结构的表示方式没有固定的格式，还没有形成标准，但基本原理是相同的。

——拓扑元素

矢量数据可抽象为点(结点)、线(链、弧段、边)、面(多边形)三种要素，即称为拓扑元素。

点(结点)——孤立点、线的端点、面的首尾点、链的连接点等。

线(链、弧段、边)——两结点间的有序弧段。

面(多边形)——若干条链构成的闭合多边形。

——最基本的拓扑关系

最基本的拓扑关系是关联和邻接。

关联——不同拓扑元素之间的关系。如结点与链，链与多边形等。

邻接——相同拓扑元素之间的关系。如结点与结点，链与链，面与面等。邻接关系是借助于不同类型的拓扑元素描述的，如面通过链面邻接。

在GIS的分析和应用功能中，还可能用到其他拓扑关系，如：

包含关系——面与其他拓扑元素之间的关系。如果点、线、面在该面内，则称为被该面包含。如某省包含的湖泊、河流等。

几何关系——拓扑元素之间的距离关系。如拓扑元素之间距离不超过某一半径的关系。

层次关系——相同拓扑元素之间的等级关系。如国家由省(自治区、直辖市)组成，省(自治区、直辖市)由县组成等。

——拓扑关系的表示

拓扑数据结构的关键是拓扑关系的表示，而几何数据的表示可参照矢量数据的简单数据结构。在目前的GIS中，主要表示基本的拓扑关系，而且表示方法不尽相同。下面举一表示矢量数据拓扑关系的例子(图2—1)。

在图2—1中，有面A、B、C、D、E、F，有链L_1、L_2、L_3、L_4、L_5、L_6、L_7、L_8、L_9、L_{10}、L_{11}、L_{12}、L_{13}，有结点P_1、P_2、P_3、P_4、P_5、P_6、P_7、P_8、

P_9。用图 2—2 来表示拓扑关系，则拓扑关系表示为：

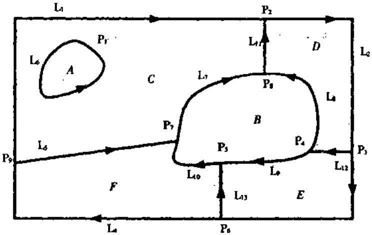


图 2—1 矢量数据拓扑关系举例

面——链关系：

面	构成面的链

链——结点关系：

链	链两端点的结点

结点——链关系：

结点	通过该结点的链

链——面关系：

链	左面	右面

图 2—2 拓扑关系的表示

面——链关系：(其中“－”号表示边的方向与构成面的方向相反。)

A	L_6
B	L_7、$-L_8$、L_9、L_{10}
C	L_1、$-L_{11}$、$-L_7$、$-L_5$
D	L_{11}、L_2、L_{12}、L_8
E	L_{13}、$-L_9$、$-L_{12}$、L_3
F	L_4、L_5、$-L_{10}$、$-L_{13}$

链——结点关系：

L_1	P_9、P_2
L_2	P_2、P_3
L_3	P_3、P_6
L_4	P_6、P_9
⋮	⋮

结点——链关系：

P_1	L_6
P_2	L_1、L_{11}、L_2
P_3	L_2、L_{12}、L_3
P_4	L_{12}、L_8、L_9
⋮	⋮

链——面关系(其中 O 为制图区域外部的多边形，常称为包络多边形)：

L_1	O	C
L_2	O	D
L_3	O	E
L_4	O	F
L_5	C	F
⋮	⋮	

§2.2 栅格数据结构

1. 基本概念

栅格数据结构——以规则的像元阵列来表示空间地物或现象

的分布的数据结构，其阵列中的每个数据表示地物或现象的属性特征。换句话说，栅格数据结构就是像元阵列，用每个像元的行列号确定位置，用每个像元的值表示实体的类型、等级等的属性编码。

对于栅格数据结构：

点实体——表示为一个像元，如图 2—3(1)。

线实体——表示为在一定方向上连接成串的相邻像元的集合，如图 2—3(2)。

面实体——表示为聚集在一起的相邻像元的集合，如图 2—3(3)。

(1)

0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 2 0 0
0 0 3 0 0 0 0 0
0 0 0 0 0 0 0 0

(2)

0 0 0 0 0 0 0 0
0 3 3 0 0 0 0 0
0 0 0 3 0 0 0 0
0 0 0 0 3 0 0 0
0 0 0 0 3 3 0 0
0 0 0 0 0 3 0 0

(3)

0 0 0 0 4 4 4 0
0 0 0 4 4 4 4 0
0 0 0 4 4 0 0 0
0 0 0 4 5 5 0 0
0 0 0 0 5 5 5 0
0 0 0 5 5 5 0 0

图 2—3 栅格数据结构表示点、线、面

栅格数据表示的是二维表面上的地理数据的离散化数值。在栅格数据中，地表被分割为相互邻接、规则排列的矩形方块(有时也可以是三角形、六边形等)，每个地块与一个像元相对应。因此，栅格数据的比例尺就是栅格(像元)的大小与地表相应单元的大小之比，当像元所表示的面积较大时，对长度、面积等的量测有较大影响。每个像元的属性是地表相应区域内地理数据的近似值，因而有可能产生属性方面的偏差。如 Landsat MSS 卫星影像的单个像元对应着地表 $79\times79m^2$ 的矩形区域，影像记录的光谱数据是每个像元所对应的地表区域内所有地物光谱辐射的总和。

栅格数据记录的是属性数据本身，而位置数据可以由属性数据对应的行列号转换为相应的坐标。栅格数据的阵列方式很容易为计算机存贮和操作，不仅很直观，而且易于维护和修改。由于栅

格数据的数据结构简单,定位存取性能好,因而在 GIS 中可与影像数据和 DEM 数据进行联合空间分析。

栅格数据的获取方式通常有:

(1)来自于遥感数据。通过遥感手段获得的数字图像就是一种栅格数据。它是遥感传感器在某个特定的时间、对一个区域地面景象的辐射和反射能量的扫描抽样,并按不同的光谱段分光并量化后,以数字形式记录下来的像素值序列。

(2)来自于对图片的扫描。通过扫描仪对地图或其他图件的扫描,可把资料转换为栅格形式的数据。

(3)由矢量数据转换而来。通过运用矢量数据栅格化技术,把矢量数据转换成栅格数据。这种情况通常是为了有利于 GIS 中的某些操作,如叠加分析等,或者是为了有利于输出。

(4)由手工方法获取。在专题图上均匀划分网格,逐个网格地确定其属性代码的值,最后形成栅格数据文件。为了保证数据的质量,在确定栅格数据中某一像元点的代码时,通常采用的方法有:

· 中心归属法——每个栅格单元的值由该栅格的中心点所在的面域的属性来确定。

· 长度占优法——每个栅格单元的值由该栅格中线段最长的实体的属性来确定。

· 面积占优法——每个栅格单元的值由该栅格中单元面积最大的实体的属性来确定。

· 重要性法——根据栅格内不同地物的重要性,选取最重要的地物的类型作为栅格单元的属性值。这种方法适用于具有特殊意义面面积较小的实体要素。

2. 栅格数据结构的表示

(1)直接栅格编码。将栅格数据看作一个数据矩阵,逐行(或逐列)记录代码,可以每行都从左到右记录,也可以奇数行从左到右,偶数行从右到左。如图 2—4 的栅格数据可存储记录为:AAAAABBBAABBAABB。

A	A	A	A
A	B	B	B
A	A	B	B
A	A	B	B

图2—4 栅格数据举例

这种记录栅格数据的文件常称为栅格文件，且常在文件头中存有该栅格数据的长和宽，即行数和列数。这样，具体的像元值就可连续存储了。其特点是处理方便，但没有压缩。

(2)游程长度(行程)编码。地理数据往往有较强的相关性，也就是说相邻像元的值往往是相同的。游程长度编码的基本思想是：按行扫描，将相邻等值的像元合并，并记录代码的重复个数。

对于图2—4，其编码为A4 A1 B3 A2 B2 A2 B2。若在行与行之间不间断地连续编码，则为A5 B3 A2 B2 A2 B2。

对于游程长度编码，区域越大，数据的相关性越强，则压缩越大。其特点是，压缩效率较高，叠加、合并等运算简单，编码和解码运算快。

(3)四叉树编码。四叉树编码是最有效的栅格数据压缩编码方法之一，在GIS中有广泛的应用。其基本思路为：将$2^n \times 2^n$像元组成的图像(不足的用背景补上)所构成的二维平面按四个象限进行递归分割，直到子象限的数值单调为止，最后得到一棵四分叉的倒向树，该树最高为n级。对于图2—4所构成的图像，可用四叉树编码法得到如下图所示的四叉树(图2—5)。

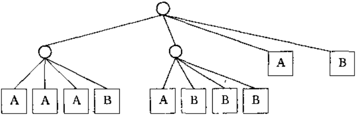


图2—5 栅格数据的四叉树编码

常规四叉树除了要记录叶结点外，还要记录中间结点，结点之间的联系靠指针。因此，为了记录常规四叉树，通常每个结点需要 6 个变量，即父结点指针、四个子结点的指针和本结点的属性值。

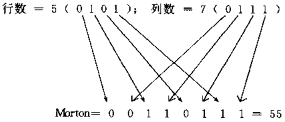
结点所代表的图像块的大小可由结点所在的层次决定，层次数由从父结点移到根结点的次数来确定。结点所代表的图像块的位置需要从根节点开始逐步推算下来。因而常规四叉树是比较复杂的。为了解决四叉树的推算问题，提出了一些不同的编码。下面介绍最常用的线性四叉树编码。

线性四叉树编码的基本思想是：不需记录中间结点和使用指针，仅记录叶结点，并用地址码表示叶结点的位置。

线性四叉树有四进制和十进制两种，下面介绍的是通常使用的十进制四叉树编码。十进制四叉树的地址码又称 Morton 码。

为了得到线性四叉树的地址码，首先将二维栅格数据的行列号转化为二进制数，然后交叉放入 Morton 码中，即为线性四叉树的地址码。

例如，对于第 5 行、第 7 列的 Moton 码为：



这样，在一个 $2^n \times 2^n$ 的图像中，每个像元点都给出一个 Morton 码，当 n=3 时即为(图 2—6)：

行＼列	0	1	2	3	4	5	6	7
0	0	1	4	5	16	17	20	21
1	2	3	6	7	18	19	22	23
2	8	9	12	13	24	25	28	29
3	10	11	14	15	26	27	30	31
4	32	33	36	37	48	49	52	53
5	34	35	38	39	50	51	54	55
6	40	41	44	45	56	57	60	61
7	42	43	46	47	58	59	62	63

图 2—6 线性四叉树的 Morton 码

这样就可将用行列表示的二维图像，用 Morton 码写成一维数据，通过 Morton 码就可知道像元的位置。

把一幅 $2^n \times 2^n$ 的图像压缩成线性四叉树的过程为：

①按 Morton 码把图像读入一维数组。

②相邻的四个像元比较，一致的合并，只记录第一个像元的 Morton 码。

③比较所形成的大块，相同的再合并，直到不能合并为止。

对用上述线性四叉树的编码方法所形成的数据还可进一步用游程长度编码压缩。压缩时只记录第一个像元的 Morton 码。

对于图 2--4 所示图像的 Morton 码为：

A_0	A_1	A_4	A_5
A_0	B_3	B_6	B_7
A_0	A_9	B_{12}	B_{13}
A_{10}	A_{11}	B_{14}	B_{15}

图 2—7 Morton 码举例

该图中像元值的右下脚标为 Morton 码，则压缩处理过程为：

①按 Morton 码读入一维数组，上一行为 Morton 码，下一行为像元值。

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

A A A B A B B B A A A A B B B B

②四相邻像元合并，只记录第一个像元的 Morton 码。

0 1 2 3 4 5 6 7 8 12

A A A B A A B B A B

③由于不能进一步合并，则用游程长度编码压缩。

0 3 4 6 8 12

A B A B A B

解码时，根据 Morton 码就可知道像元在图像中的位置（相对左上角），本 Morton 码和下一个 Morton 码之差即为像元个数。知道了像元的个数和像元的位置就可恢复出图像了。

线性四叉树编码的优点是：压缩效率高，压缩和解压缩比较方便，阵列各部分的分辨率可不同，既可精确地表示图形结构，又可减少存贮量，易于进行大部分图形操作和运算。缺点是：不利于形状分析和模式识别，即具有变换不定性，如同一形状和大小的多边形可得出完全不同的四叉树结构。

§2.3 矢量数据结构与栅格数据结构的比较

矢量数据结构可具体分为点、线、面，可以构成现实世界中各种复杂的实体，当问题可描述成线或边界时，特别有效。矢量数据的结构紧凑，冗余度低，并具有空间实体的拓扑信息，容易定义和操作单个空间实体，便于网络分析。矢量数据的输出质量好、精度高。

矢量数据结构的复杂性，导致了操作和算法的复杂化，作为一种基于线和边界的编码方法，不能有效地支持影像代数运算，如不能有效地进行点集的集合运算（如叠加），运算效率低而复杂。由于矢量数据结构的存贮比较复杂，导致空间实体的查询十分费时，需要逐点、逐线、逐面地查询。矢量数据和栅格数据表示的影像数据

不能直接运算(如联合查询和空间分析),交互时必须进行矢量和栅格转换。矢量数据与DEM(数字高程模型)的交互是通过等高线来实现的,不能与DEM直接进行联合空间分析。

栅格数据结构是通过空间点的密集而规则的排列表示整体的空间现象的。其数据结构简单,定位存取性能好,可以与影像和DEM数据进行联合空间分析,数据共享容易实现,对栅格数据的操作比较容易。

栅格数据的数据量与格网间距的平方成反比,较高的几何精度的代价是数据量的极大增加。因为只使用行和列来作为空间实体的位置标识,故难以获取空间实体的拓扑信息,难以进行网络分析等操作。栅格数据结构不是面向实体的,各种实体往往是叠加在一起反映出来的,因而难以识别和分离。对点实体的识别需要采用匹配技术,对线实体的识别需采用边缘检测技术,对面实体的识别则需采用影像分类技术,这些技术不仅费时,而且不能保证完全正确。

通过以上的分析可以看出,矢量数据结构和栅格数据结构的优缺点是互补的,为了有效地实现GIS中的各项功能(如与遥感数据的结合,有效的空间分析等)需要同时使用两种数据结构,并在GIS中实现两种数据结构的高效转换。

第三章　GIS的空间数据库和数据模型

§3.1　空间数据库基础

1. 数据库的概念

数据库技术产生于20世纪60年代末期，是计算机领域中最重要的技术之一，是一种较理想的数据管理技术。所谓数据管理是指对数据的组织、存贮、检索和维护。用计算机进行数据管理大体经历了三个阶段：即人工管理阶段、文件系统阶段和数据库阶段。

由于文件系统存在着文件之间无联系、数据不完全独立于程序，只在文件级共享数据等缺陷，因而数据库就发展了起来。

数据库可以看作是与现实世界有一定相似性的模型，是认识世界的基础，是集中、统一地存储和管理某个领域信息的系统，它根据数据间的自然联系而构成，数据较少冗余，且具有较高的数据独立性，能为多种应用服务。

数据库作为一个复杂的系统，由以下三个基本部分构成：

(1)数据集。一个结构化的相关数据的集合体，包括数据本身和数据间的联系。数据集独立于应用程序而存在，是数据库的核心和管理对象。

(2)物理存储介质。指计算机的外存储器和内存储器。前者存储数据；后者存储操作系统和数据库管理系统，并有一定数量的缓冲区，用于数据处理，以减少内外存交换次数，提高数据存取效率。

(3)数据库软件。其核心是数据库管理系统(DBMS)。主要任务是对数据库进行管理和维护。具有对数据进行定义、描述、操作

和维护等功能，接受并完成用户程序和终端命令对数据库的请求，负责数据库的安全。

2. 数据库的主要特征

数据库方法与文件系统方法相比，具有更强的数据管理能力。数据库具有如下主要特征：

(1)数据集中控制。在文件管理方法中，文件是分散的，每个用户或每种处理都有各自的文件，这些文件之间一般是没有联系的，因此，不能按照统一的方法来控制、维护和管理。而数据库则很好地克服了这一缺点，可以集中控制、维护和管理有关数据。

(2)数据独立。数据库中的数据独立于应用程序，包括数据的物理独立性和逻辑独立性，给数据库的使用、调整、优化和进一步扩充提供了方便，提高了数据库应用系统的稳定性。

(3)数据共享。数据库中的数据可以供多个用户使用，每个用户只与库中的一部分数据发生联系；用户数据可以重叠，用户可以同时存取数据而互不影响，大大提高了数据库的使用效率。

(4)减少数据冗余。数据库中的数据不是面向应用，而是面向系统。数据统一定义、组织和存储，集中管理，避免了不必要的数据冗余，也提高了数据的一致性。

(5)数据结构化。整个数据库按一定的结构形式构成，数据在记录内部和记录类型之间相互关联，用户可通过不同的路径存取数据。

(6)统一的数据保护功能。在多用户共享数据资源的情况下，对用户使用数据有严格的检查，对数据库规定密码或存取权限，拒绝非法用户进入数据库，以确保数据的安全性、一致性和并发控制。

3. 数据库的系统结构

数据库系统的结构，一般划分为三个层次，它们是概念模式、外模式和内模式(图 3--1)。

(1)概念模式。亦称模式，是数据库的总框架，描述数据库中关

于目标存储的逻辑结构、基本操作、目标——目标及目标——操作的关系和依赖性，以及对数据的安全性、完整性等方面的定义。所有数据都按这一模式进行装配。概念模式由概念模式描述语言DDL来进行描述。

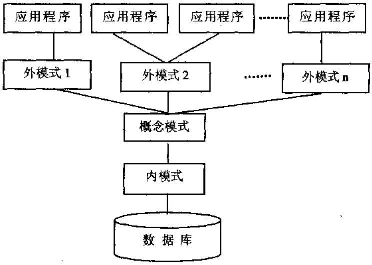


图 3—1 数据库系统的结构

(2)外模式。亦称子模式。是数据库用户的数据视图。它属于概念模式的一部分，描述用户数据的结构、类型、长度等。所有的应用程序都是根据外模式中对数据的描述而不是根据概念模式中对数据的描述而编写的。在一个外模式中可以编写多个应用程序，但一个应用程序只能对应一个外模式。根据应用的不同，一个概念模式可以对应多个外模式，外模式可以互相覆盖。外模式由外模式描述语言SDDL进行具体描述。

(3)内模式。亦称存储模式。是对数据库在物理存储器上具体实现的描述。它规定数据在存储介质上的物理组织方式、记录寻址技术，定义物理存储块的大小，溢出处理方法等。与概念模式相对应。内模式由数据存储描述语言DSDL进行描述。

数据库系统的三级模式结构将数据库的全局逻辑结构同用户的局部逻辑结构和物理存储结构区分开来，给数据库的组织和使用带来了方便。不同的用户可以有各自的数据视图，所有用户的数据视图集中在一起统一组织，消除冗余数据，得到全局数据视图。用存储描述语言来定义和描述全局数据视图数据，并将数据存储在物理介质上。这中间进行了两次映象。一次是外模式与概念模式之间的映象，定义了它们之间的对应关系，保证了数据的逻辑独立性；另一次是概念模式与内模式之间的映象，定义了数据的逻辑结构和物理存储之间的对应关系，使全局逻辑数据独立于物理数据，保证了数据的物理独立性。

在数据库系统中，用户所看到的数据和计算机中存放的数据是两回事。用户所理解的是用户自己理解的实体、实体的属性和实体间的关系，即用户只看到了数据库的外层。

4. 数据模型

在数据库系统中，现实世界中的事物及联系是用数据模型来描述的，数据库中各种操作功能的实现是基于不同的数据模型的，因而数据库的核心问题是模型问题。数据模型是数据库中对数据的逻辑组织形式的描述。

数据模型是对现实世界部分现象的抽象，它描述了数据的基本结构及其相互之间的关系和在数据上的各种操作，是数据库系统中关于数据内容和数据间联系的逻辑组织的形式表示，以抽象的形式描述和反映一个部门或系统的业务活动和信息流程。

选择与建立数据模型的目的是用最佳的方式反映本部门的业务对象及信息流程和以最佳的方式为用户提供访问数据库的逻辑接口。

数据模型的三要素：数据结构、数据操作和数据的约束条件。

数据模型可以分为三种类型：一种是独立于计算机之外的，如实体——关系模型、语义数据模型等，它们不涉及信息在计算机中如何表示，常称概念模型；另一种模型是直接面向计算机的，它们

以记录为单位构造数据模型，如数据库中常用的层次模型、网状模型和关系模型等，通常称为逻辑模型；还有一种称为物理模型，是概念模型和逻辑模型在计算机内部的存储和操作形式。在数据模型中常用到下列概念：

· 实体(Entity)：实体是指现实世界中客观存在的，并可相互区别的事物。实体可以指个体，也可以指总体，即个体的集合。

· 属性(Attribute)：实体所具有的某一特性

· 码(Key)：唯一标识实体的属性集

· 域(Domain)：属性的取值范围

· 实体型(Entity Type)：具有相同属性的实体具有共同的特征和性质，用实体名和属性名集合表示

· 实体集(Entity Set)：同型实体的集合

· 联系：一是实体内部的联系；二是实体型之间的联系(1:1 、1:n、m:n)。

数据模型是数据特征的抽象，它不是描述个别数据，而是描述数据的共性。严格地说，一个数据库的数据模型应能描述数据的以下特征：

①静态特性。包括实体和实体具有的特性、实体间的联系等，通过构造基本数据结构类型来实现。

②动态特性。即现实世界中的实体及实体间的不断发展变化，通过对数据库的检索、插入、删除和修改等操作来实现。

③数据间的相互制约与依存关系。通过一组完整性规则来实现。

由此可见，一个数据模型实际上给出了在计算机系统中描述现实世界的信息结构及其变化的一种抽象方法。数据模型不同，描述和实现的方法也不相同，相应的支持软件——数据库管理系统也就不同。

数据模型反映了现实世界中的实体之间的各种联系。实体间的联系有两类：一类是实体内部属性间的联系；另一类是实体与实

体之间的联系。

实体与实体之间的联系是错综复杂的，可以分为以下三种：

①一对一的联系。这是最简单的一种实体之间的联系，它表示两个实体集中的个体间存在的一对一的联系。记为 1:1。

②一对多的联系。这是实体间存在的较普遍的一种联系，表示一种实体集 E_1 中的每个实体与另一实体集 E_2 中的多个实体间存在的联系；反之，E_2 中的每个实体都至多与 E_1 中的一个实体发生联系。记为 1:m。

③多对多的联系。这是实体间存在的较为普遍的一种联系，表示多个实体集之间的多对多的联系。其中，一个实体集中的任何一个实体与另一个实体集中的实体间存在一对多的联系；反之亦然。记为 m:n。

5. 数据库管理系统(DBMS)

数据库管理系统(DBMS)是处理数据库中数据存取和各种管理控制的软件。它是数据库系统的核心，应用程序对数据库的操作全部通过 DBMS 进行。数据库管理系统(DBMS)通常具有如下功能：

(1)数据库定义功能。用数据库的数据描述语言 DDL 来定义概念模式、外模式和内模式，也就是说，具有给出数据库框架的功能。如定义数据库的逻辑结构、数据库的框架结构、定义数据项、建立记录类型、定义记录间的关系、指定安全控制要求等。

(2)数据库管理功能。指对数据进行更新、存取等控制功能。通常提供有数据操作语言来作为用户和数据库之间的接口。常用的数据库管理功能如：从数据库中检索出满足条件的数据、向数据库中插入数据、删除数据、修改数据、进行控制操作(如并发控制)等。

(3)数据库维护功能。数据库需要维护，通常包括如下工作：

· 改善系统的性能：及时掌握数据库的性能变化，性能下降时应进行干预，如对数据进行重新整理和组织。

· 受损后的复原：一方面应能防止各种伤害数据库的行为，另

一方面当数据库受损后，应具有复原的手段。

·用户管理：对用户应统一管理，分配使用权限，防止非法使用。

·拓宽数据库用户的要求：根据用户要求，修改数据模式，根据新模式重新组织数据。

(4)通讯功能：应具有与操作系统的接口、与各种语言的接口、与其他数据库通讯等能力。

6. 空间数据库与地理实体

地理信息通过数据采集和编辑以后，送入到计算机的外存设备。对于海量的空间数据，采用文件系统的方法来管理肯定不行，必须采用数据库技术进行管理。因此，空间数据库成了 GIS 研究的重要课题。空间数据库是地理空间数据的集合，是一种与现实的地理世界保持一定相似性的实体模型。

地理实体——空间数据库中的实体，是一种在现实世界中不能再划分为同类现象的现象。例如城市可看成一个地理实体，并可划分成若干部分，但这些部分不叫城市，只能称为区、街道之类。

地理目标——实体在空间数据库中的表示。地理目标的表示方法随比例尺等情况的变化而变化，例如，对于城市这个地理实体，在小比例尺上可作为一个点目标，而在大比例尺上将作为一个面目标。地理目标在地图上是以地图符号的形式来表示的。

(1)地理实体的类型。以相同的方式表示和存储的一组类似的地理实体，可以作为地理实体的一种类型。空间数据库设计的第一步就是选择和定义所需要的实体类型。

地理实体通常分为点状实体、线状实体、面状实体和体状实体，复杂的地理实体由这些类型的实体构成。

——点状实体。点状实体是指只有特定的位置，而没有长度的实体。如：

实体点——用于代表一个实体；

注记点——用于定位注记；

内点——用于负载相应多边形的属性；

结点——表示线的终点和起点；

节点——线或弧段的内部点。

——线状实体。线状实体是指有长度的实体，如线段、边界、链、网络等，并且有如下特性：

长度——从起点到终点的总长；

曲率——用于表示线状实体的弯曲程度，如道路拐弯处；

方向——如水流的方向等。

——面状实体。面状实体也称多边形、区域等，是对湖泊、岛屿、地块等一类现象的描述，通常有如下空间特征：

面积——面状实体所占有的范围的大小；

周长——面状实体所占有区域的周长；

独立或相邻——是独立存在，还是与其他面状实体相邻；

岛或洞——面状实体中是否有岛或洞；

重叠——面状实体之间是否有重叠。

——体状实体。体状实体用于描述三维空间中的现象与物体，它具有长度、宽度及高度等属性，通常有如下空间特征：

体积；

岛或洞；

表面积。

(2)地理目标的类型。空间数据库中实体类型的表示需要选择合适的地理目标类型，而地理目标的类型可按空间维来定义。

0维——有位置无长度的目标，如点。

1维——有长度的目标，一般由两个或多个0维目标组成，如线。

2维——有长和宽的目标，如多边形。

3维——有长、宽和高的目标，如三维立体。

(3)地理实体的描述。通常需要从如下方面对地理实体进行描述：

编码——用于区别不同的实体，有时同一个实体在不同的时间具有不同的编码，如上行和下行的火车。编码通常包括分类码和识别码。分类码标识实体所属的类别，识别码对每个实体进行标识，是唯一的，用于区别不同的实体。

位置——通常用坐标值的形式（或其他方式）给出实体的空间位置。

类型——指明该地理实体属于哪一种实体类型，或由哪些实体类型组成。

行为——指明该地理实体可以具有哪些行为和功能。

属性——指明该地理实体所对应的非空间信息，如道路的宽度、路面质量、车流量、交通规则等。

说明——用于说明实体数据的来源、质量等相关的信息。

关系——与其他实体的关系信息。

（4）地理实体时间维的描述。如果只是地理实体的属性数据在变化，那么，可以把不同时间的属性数据均记录下来，作为该地理实体的属性数据。例如在处理统计区域的人口数时，区域的空间位置不变，只要把新的人口数及对应的时间加入到属性数据表中即可。

当地理实体的空间位置随时间变化时，如政区界线的变化、地块的合并与重新划分等，这时必须把地理实体的空间特征的变化也记录下来，如记录实体的增加、删除、改变、移动、合并等，同时对实体进行时间标记。

7. 空间数据的分层

空间数据可按某种属性特征形成一个数据层，通常称为图层。图层是描述某一地理区域的某一（有时也可以是多个）属性特征的数据集。因此，某一区域的空间数据可以看成是若干图层的集合。

原则上讲图层的数量是无限制的，但实际上要受 GIS 数据结构、计算机存储空间等的限制。通常按以下方法对空间数据进行分层：

(1)按专题分层。每个图层对应一个专题，包含某一种或某一类数据。如地貌层、水系层、道路层、居民地层等。对于不同的研究目的，空间数据可以根据不同的专题分成不同的数据层。

(2)按时间序列分层。即把不同时间或不同时期的数据作为一个数据层。

空间数据分层的目的主要是为了便于空间数据的管理、查询、显示、分析等。当空间数据分为若干数据层后，对所有空间数据的管理就简化为对各数据层的管理，而一个数据层的数据结构往往比较单一，数据量也相对较小，管理起来就相对简单；而对分层的空间数据进行查询时，不需要对所有空间数据进行查询，只需要对某一层空间数据进行查询即可，因而可加快查询速度；分层后的空间数据由于可以任意选择需要显示的图层，因而增加了图形显示的灵活性；对不同数据层进行叠加，可进行各种目的的空间分析。

§3.2 GIS的传统空间数据模型

1. 层次数据模型

层次模型是一种树结构模型，它把数据按自然的层次关系组织起来，以反映数据之间的隶属关系。层次模型是数据库技术中发展最早、技术上比较成熟的一种数据模型。它的特点是空间数据组织成有向有序的树结构，也叫树形结构。结构中的结点代表数据记录，连线描述位于不同结点数据间的从属关系(一对多的关系)。

由树的定义知，一棵树有且仅有一个无双亲结点的称为根的结点；其余结点有且仅有一个双亲结点，它们可分为m(m≥0)个互不相交的有限集，其中每一个集合本身又是一棵树，将其称为子树。

图3—2表示地理实体E及其空间要素，图3—3是图3—2所示空间关系所构成的层次模型。这是一棵有向有序树，结点表示不同层次的地理要素，连线描述地理要素之间的从属关系。结点从属

于(构成)有向边,有向边从属于(构成)多边形,多边形从属于(构成)实体 E。

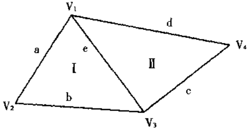


图 3—2 实体 E 及其空间要素

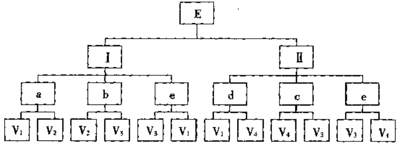


图 3—3 层次模型

2. 网状数据模型

网状模型将数据组织成有向图结构,图中的结点代表数据记录,连线描述不同结点数据间的联系。这种数据模型的基本特征是:结点数据之间没有明确的从属关系,一个结点可与其他多个结点建立联系,即结点之间的联系是任意的,任何两个结点之间都能发生联系,可表示多对多的关系。图 3—2 所示实体 E 及其空间要素的网状模型如图 3—4 所示。

采用网状模型最典型的数据库系统是 DBTG 系统,它是 1969 年由一个美国标准化组织 CODASYL 委员会的数据库任务组提出的报告中首次推出的,这个报告是一个网状模型的数据描述语

言和数据操纵语言规范化的文本,虽不是具体计算机的软件系统,但对网状数据库系统的研究和发展起了重要作用,现有的网状数据库系统大都是建立在 DBTG 模型上的。

层次数据模型和网状数据模型的区别:

· 层次模型中从子女到双亲的联系是唯一的,而网状模型则可以不唯一;

· 层次模型不允许有复合链,而网状模型则允许;

层次数据模型和网状数据模型的缺陷:

· 要求应用程序员必须熟悉面向磁盘的优化技术和数据库的物理组织,对于每次特定的数据查询,必须编出十分复杂的查询应用程序。

· 一旦有新类型的数据加入,将会导致数据库结构的变化,在这种情况下,通常应用程序需要重写。

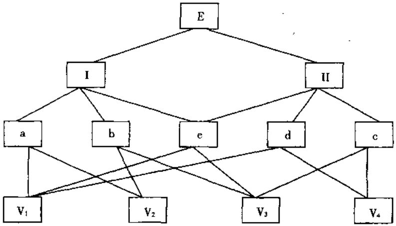


图 3-4 网状模型

3. 关系数据模型

(1)关系模型的逻辑数据结构是 IBM 公司的 E · F · Codd 提出来的。他从 1970 年起发表了多篇关于关系模型的论文,奠定了

关系数据库的理论基础。由于关系数据库结构简单，操作方便，有坚实的理论基础，所以发展很快，80 年代以后推出的数据库管理系统几乎都是关系型的。

在关系模型中，数据的逻辑结构为满足一定条件的二维表，表具有固定的列数和任意的行数，在数学上称为“关系”。二维表是同类实体的各种属性的集合，每个实体对应于表中的一行，在关系中称为元组，相当于通常的一个记录；表中的列表示属性，称为域，相当于通常记录中的一个数据项。若二维表中有 n 个域，则每一行叫做一个 n 元组，这样的关系称为 n 度(元)关系。表的行对应于对象的实例，各个表的行列交点就用来存贮简单值。满足一定条件的规范化关系的集合，就构成了关系模型。

关系模型可以简单、灵活地表示各种实体及其关系，其数据描述具有较强的一致性和独立性。在关系数据库系统中，对数据的操作是通过关系代数实现的，具有严格的数学基础。

关系数据模型要求每一属性都必须赋给一个域，一个域就是一些正确值的集合，域可允许更多的语义检查，表中的每一个值都必须属于其属性的域或者为空。

RDBMS 使用各种手段来加速表的访问，这些协调手段是透明的，只在读写表的命令中是不可见的。由 RDBMS 来决定在处理查询的过程使用协调信息的时间，如果需要则自动执行协调。每当修改相应表时，RDBMS 自动修改协调信息。索引、hashing 及排序都是常用的协调技术。

(2)表的操作符。SQL(Structured Query Language)语言是 1974 年由 Boyce 和 Chamberlin 提出的，在 IBM 公司 San Jose Research Laboratory 研制的 System R 上实现了这种语言，1986 年 10 月，美国国家标准局(ANSI)批准 SQL 作为关系数据库语言的国家标准，此后不久，国际标准化组织(ISO)也作出了同样的决定。

SQL 提供了表的操作符，它的功能包括查询(Query)、操纵

(Manipulation)、定义(Defifition)和控制(Control)四个方面,是一个综合的、通用的、功能强大的关系数据库语言。

(3)表的完整性规则。实体完整性:每个表有一个明确的主关键字,主关键字是一个或者多个属性的集合;关联完整性:要求RDBMS保持其外来关键字与它相应的主关键字一致,外来关键字是一个表的主关键字但它又嵌入另一个表中。

此外,还有两个概念也需要给予一定的解释。

· 视图(View):一个虚表,即视图所对应的数据并不是存储在视图中,而是存在导出它的一个或多个表中。

· 范式:用来避免表修改操作的逻辑不一致性的一些规则集,范式有多种层次,高级范式可以对低级范式增加约束条件。在1971～1972年,Codd系统地提出了第一范式(1NF),第二范式(2NF),第三范式(3NF)的概念,1974年,他和Boyce又共同提出了BCNF,后来又有人提出了4NF和5NF。其中

1NF:如果一关系模式,它的每一个分量是不可分的数据项,则此关系模式为1NF;

2NF:若关系模式R∈1NF,且每个非主属性完全函数依赖于码,则R∈2NF;

3NF:若关系模式R(U,F)∈1NF,不存在非主属性对码的传递依赖,则R∈3NF;

BCNF:关系模式R(U,F)∈1NF,若X－>Y,Y⊄X时,X必含有码,则说R∈BCNF;

第一范式的概念限制了系统对复杂对象的表示,从而使关系数据库对非传统的应用领域束手无策,为了使关系数据库适应现实的需求,有人提出用NF2(Non－First－Normal－Form)模型,这种模型允许关系作为属性值。但仅仅在理论上有些价值,而对于实用,若引入NF2,则不仅要修改数据模型,而且要较大地改动RDBMS中的其他许多部分,如存贮结构、存取方法等等,这在实现上很困难。

(4)用关系模型表示地理实体。图 3—2 所示地理实体 E 与空间要素可用关系模型构成如下关系：

地理实体—多边形关系：E(Ⅰ,Ⅱ)

多边形—— 多边形关系：Ⅰ(a,b,e)

Ⅱ(e,c,d)

边——结点关系：a(V1,V2)

b(V2,V3)

c(V3,V4)

d(V1,V4)

e(V1,V3)

上述关系的二维表实现如图 3—5 所示。

E	Ⅰ
E	Ⅱ

Ⅰ	a	b	e
Ⅱ	e	c	d

a	V_1	V_2
b	V_2	V_3
e	V_3	V_1
d	V_1	V_4
c	V_4	V_3

图 3-5 关系模型

(5)关系模型的进一步发展。RDBMS 在目前的现状是理论和技术都非常完善和成熟，商品化的产品大量涌向市场，基本上占据了目前的数据库市场中传统应用领域的全部和 90%以上的非传统应用领域，典型的代表有：DB2、INGRES、Oracle、SyBase、Xbase、FoxPro、Rdb/VMS、Informix 等等。RDBMS 的发展主要体现在以下方面：

——非结构化大型对象(LOB－Large Object)的引进。在关

系模型中附加非结构化内容，用于存贮大量变长字符串和二进制数据，是 DEC 公司于 1981 年就引进的概念，只是当时并未流行。随着多媒体技术的推广和应用，需要对诸如文本数据、图像数据、静止的视频片段、全动感的视频信息、声音波形等等这类大量的非结构化数据进行存储和管理，于是在目前一些大型的 RDBMS(如 DB2、Oracle 等)中，增加了大型对象这种数据类型。

——分布式数据库。20 世纪 70 年代中期以来，由于计算机网络通信的迅速发展以及在地理上分散的公司、团体和组织对于数据库更为广泛的应用需求，在集中式数据库系统成熟技术的基础上产生和发展了分布式数据库系统，分布式数据库系统是数据库技术和网络技术两者相互渗透和有机结合的结果。目前，在该领域里，一些工作已经取得了显著的成果；许多基本问题已得到解决；一批原型系统已经研究成功并获得了许多经验；一些商品化的产品正在研制或已经推出。如 Oracle、Informix、Sybase 等等。

——对象特性的融入。为了迎接面向对象数据库在非传统应用领域对关系数据库提出的挑战，确保关系数据库在数据库领域的霸主地位，关系数据库巨头 Oracle、Sybase、DB2、Informix 都在积极努力，使自己的关系型产品增加一些面向对象的成分。目前已经有推向市场的这类产品存在，如 Oracle8、DB2 等。

§3.3 GIS 的对象数据模型

1. 传统数据模型存在的主要问题

已于前述，目前非空间数据最主要的数据模型是层次模型、网状模型和关系模型。这里，我们分别介绍它们用于 GIS 空间数据库的局限性。

(1)层次模型用于 GIS 空间数据库的局限性。层次模型反映了地理世界中实体之间的层次关系，在描述地理世界中自然的层次结构关系时简单、直观，易于理解，并在一定程度上支持数据的

重构。它用于GIS空间数据库存在的主要问题是：

——很难描述复杂的地理实体之间的联系，描述多对多的关系时导致物理存储上的冗余；

——对任何对象的查询都必须从层次结构的根结点开始，低层次对象的查询效率很低，很难进行反向查询；

——数据独立性较差，数据更新涉及许多指针，插入和删除操作比较复杂，父结点的删除意味着其下层所有子结点均被删除；

——层次命令具有过程性质，要求用户了解数据的物理结构，并在数据操纵命令中显式地给出数据的存取路径；

——基本不具备演绎功能和操作代数基础。

(2)网状模型用于GIS空间数据库的局限性。网状模型是层次模型的一般形式，反映了地理世界中常见的多对多关系，在一定程度上支持数据的重构，具有一定的数据独立和数据共享特性，且运行效率较高。用于GIS空间数据库的主要问题如下：

——由于网状结构的复杂性，增加了用户查询的定位困难，要求用户熟悉数据的逻辑结构，知道自己所处的位置；

——网状数据操作命令具有过程性质，存在与层次模型相同的问题；

——不直接支持对于层次结构的表达；

——基本不具备演绎功能和操作代数基础。

(3)关系模型用于GIS空间数据库的局限性。关系模型表示各种地理实体及其间的关系，方式简单、灵活，支持数据重构；具有严格的数学基础，并与一阶逻辑理论密切相关，具有一定的演绎功能；关系操作和关系演算具有非过程特点。尽管如此，关系模型用于GIS空间数据库也还存在一些不足。主要问题是：

——无法用递归和嵌套的方式来描述复杂关系的层次和网状结构，模拟和操作复杂地理对象的能力较弱；

——用关系模型描述本身具有复杂结构和涵义的地理对象时，需对地理实体进行不自然的分解，导致存储模式、查询途径及

操作等方面均显得语义不甚合理；

——由于概念模式和存储模式的相互独立性，实现关系之间的联系需要执行系统开销较大的联接操作，运行效率不够高。

不难看出，关系模型的根本问题是不能有效地管理复杂地理对象。

2. 面向对象的概念

面向对象的基本概念是在 20 世纪 70 年代萌发出来的，它的基本做法是把系统工程中的某个模块和构件视为问题的空间的一个或一类对象。到了 80 年代，面向对象的方法得到很快发展，在系统工程、计算机、人工智能等领域获得了广泛应用。但是，在更高级的层次上和更广泛的领域内对面向对象的方法进行研究还是 90 年代的事。

(1)基本思想和基本概念。面向对象的基本思想是通过对问题领域进行自然的分割，用更接近人类通常思维的方式建立问题领域的模型，并进行结构模拟和行为模拟，从而使设计出的软件能尽可能地直接表现出问题的求解过程。因此，面向对象的方法就是以接近人类通常思维方式的思想，将客观世界的一切实体模型化为对象。每一种对象都有各自的内部状态和运动规律，不同对象之间的相互联系和相互作用就构成了各种不同的系统。

在面向对象的方法中，对象、类、方法和消息是基本的概念。

对象——含有数据和操作方法的独立模块，可以认为是数据和行为的统一体。如一个城市、一棵树均可作为地理对象。对于一个对象，应具有如下特征：

- 具有一个唯一的标识，以表明其存在的独立性；
- 具有一组描述特征的属性，以表明其在某一时刻的状态；
- 具有一组表示行为的操作方法，用以改变对象的状态。

类——共享同一属性和方法集的所有对象的集合构成类。从一组对象中抽象出公共的方法和属性，并将它们保存在一类中，是面向对象的核心内容。如河流均具有共性，如名称、长度、流域面积

等，以及相同的操作方法，如查询、计算长度、求流域面积等，因而可抽象为河流类。被抽象的对象，称为实例，如长江、黄河等。

消息——对对象进行操作的请求，是连接对象与外部世界的唯一通道。

方法——对对象的所有操作，如对对象的数据进行操作的函数、指令、例程等。

(2)面向对象的特性。面向对象方法具有抽象性、封装性、多态性等特性。

抽象是对现实世界的简明表示。形成对象的关键是抽象，对象是抽象思维的结果。抽象思维是通过概括、判断、推理来反映对象的本质，揭示对象内部联系的过程。任何一个对象都是通过抽象和概括而形成的。面向对象方法具有很强的抽象表达能力，正是因为这个缘故，可以将对象抽象成对象类，实现抽象的数据类型，允许用户定义数据类型。

封装是指将方法与数据放于一对象中，以使对数据的操作只可通过该对象本身的方法来进行。即一对象不能直接作用于另一对象的数据，对象间的通信只能通过消息来进行。对象是一个封装好的独立模块。封装是一种信息隐蔽技术，封装的目的在于将对象的使用者和对象的设计者分开，用户只能见到对象封装界面上的信息，对象内部对用户是隐蔽的。也就是说，对用户而言，只需了解这个模块是干什么的即功能是什么，至于怎么干即如何实现这些功能则是隐蔽在对象内部的。一个对象的内部状态不受外界的影响，其内部状态的改变也不影响其他对象的内部状态。封装本身即模块性，把定义模块和实现模块分开，就使得用面向对象技术开发或设计的软件的可维护性、可修改性大为改善。

多态是指同一消息被不同对象接收时，可解释为不同的含义。因此，可以发送更一般的消息，把实现的细节都留给接收消息的对象。即相同的操作可作用于多种类型的对象，并能获得不同的结果。

(3)面向对象数据模型的四种核心技术。

——分类。类是具有相同属性结构和操作方法的对象的集合，属于同一类的对象具有相同的属性结构和操作方法。分类是把一组具有相同属性结构和操作方法的对象归纳或映射为一个公共类的过程。对象和类的关系是"实例"(instance—of)的关系。

同一个类中的若干个对象，用于类中所有对象的操作都是相同的。属性结构即属性的表现形式相同，但它们具有不同的属性值。所以，在面向对象的数据库中，只需对每个类定义一组操作，供该类中的每个对象使用，而类中每一个对象的属性值要分别存储，因为每个对象的属性值是不完全相同的。例如，在面向对象的空间数据模型中，城镇建筑可分为行政区、商业区、住宅区、文化区等若干个类。以住宅区类而论，每栋住宅作为对象都有门牌号、地址、电话号码等相同的属性结构，但具体的门牌号、地址、电话号码等是各不相同的。当然，对它们的操作方法如查询等都是相同的。

——概括。概括是把几个类中某些具有部分公共特征的属性和操作方法抽象出来，形成一个更高层次、更具一般性的超类的过程。子类和超类用来表示概括的特征，表明它们之间的关系是"即是"(is—a)关系，子类是超类的一个特例。

作为构成超类的子类还可以进一步分类，一个类可能是超类的子类，同时也可能是几个子类的超类。所以，概括可能有任意多层次。例如，建筑物是住宅的超类，住宅是建筑物的子类，但如果把住宅的概括延伸到城市住宅和农村住宅，则住宅又是城市住宅和农村住宅的超类。

概括技术的采用避免了说明和存储上的大量冗余，因为住宅地址、门牌号、电话号码等是"住宅"类的实例(属性)，同时也是它的超类"建筑物"的实例(属性)。当然，这需要一种能自动地从超类的属性和操作中获取子类对象的属性和操作的机制。

——聚集。聚集是将几个不同类的对象组合成一个更高级的复合对象的过程。术语"复合对象"用来描述更高层次的对象，"部

分"或"成分"是复合对象的组成部分,"成分"与"复合对象"的关系是"部分"(parts—of)的关系,反之"复合对象"与"成分"的关系是"组成"的关系。例如,医院由医护人员、病人、门诊部、住院部、道路等聚集而成。

每个不同属性的对象是复合对象的一个部分,它们有自己的属性数据和操作方法,这些是不能为复合对象所公用的,但复合对象可以从它们那里派生得到一些信息。复合对象有自己的属性值和操作,它只从具有不同属性的对象中提取部分属性值,且一般不继承子类对象的操作。这就是说,复合对象的操作与其成分的操作是不兼容的。

——联合。联合是将同一类对象中的几个具有部分相同属性值的对象组合起来,形成一个更高水平的集合对象的过程。术语"集合对象"描述由联合而构成的更高水平的对象,有联合关系的对象称为成员,"成员"与"集合对象"的关系是"成员"(member—of)的关系。

在联合中,强调的是整个集合对象的特征,而忽略成员对象的具体细节。集合对象通过其成员对象产生集合数据结构,集合对象的操作由其成员对象的操作组成。例如,一个农场主有三个水塘,它们使用同样的养殖方法,养殖同样的水产品,由于农场主、养殖方法和养殖水产品等三个属性都相同,故可以联合成一个包含这三个属性的集合对象。

联合与概括在概念上不同。概括是对类进行抽象概括;而联合是对属于同一类的对象进行抽象联合。联合有点类似于聚集,所以在许多文献中将联合的概念附在聚集的概念中,都使用传播工具提取对象的属性值。

(4)面向对象数据模型的核心工具。

——继承。继承为面向对象方法所独有,服务于概括。在继承体系中,子类的属性和方法依赖父类的属性和方法。继承是父类定义子类,再由子类定义其子类,一直定义下去的一种工具。父类和

子类的共同属性和操作由父类定义一次，然后由其所有子类对象继承，但子类可以有不是从父类继承下来的另外的特殊属性和操作。一个系统中，对象类是各自封装的，如果没有继承这一强有力的机制，类中的属性值和操作方法就可能出现大量重复。所以，继承是一种十分有用的抽象工具，它减少了冗余数据，又能保持数据的完整性和一致性，因为对象的本质特征只定义一次，然后由其相关的所有子类对象继承。

父类的操作适用于所有的子类对象，因为每一个子类对象同时也是父类的对象。当然，专为子类定义的操作是不适用于其父类的。

继承有单重继承和多重继承之分。

1. 单重继承：指仅有一个直接父类的继承，要求每一个类最多只能有一个中间父类，这种限制意味着一个子类只能属于一个层次，而不能同时属于几个不同的层次。如图 3—6 所示，"住宅"是父类，"城市住宅"和"农村住宅"是其子类，父类"住宅"的属性（如"住宅名"）可以被它的两个子类继承，同样，给父类"住宅"定义的操作（如"进入住宅"）也适用于它的两个子类；但是，专为一个子类定义的操作如"地铁下站"，只适用于"城市住宅"。

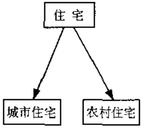


图 3-6 单重继承

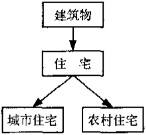


图 3—7 三个层次的单重继承

单重继承可以构成树形层次，最高父类在顶部，最特殊的子类在底部，每一类可看作一个结点，两个结点的"即是"关系可以用父类结点指向子类结点的矢量来表示，矢量的方向表示从上到下、从

一般到特殊的特点。

继承不仅可以把父类的特征传给中间子类，还可以向下传给中间子类的子类。图 3—7 是有三个层次的继承体系。“建筑物”的特征(如“户主”、“地址”等)可以传给中间子类“住宅”，也可以传给中间子类的子类“城市住宅”和“农村住宅”。

2. 多重继承：允许子类有多于一个的直接父类的继承。严格的层次结构是一种理想的模型，对现实的空间数据常常不适用。多重继承允许几个父类的属性和操作传给一个子类，这就不是层次结构。

GIS 中经常遇到多重继承问题。图 3—8 是两个不同的体系形成的多重继承的例子。一个体系为交通运输线，另一个体系为水系。运河具有人工交通运输线和河流等两个父类特性，通航河流也有自然交通运输线和河流等两个父类的特性。

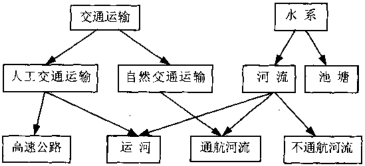


图 3—8 多重继承

——传播。传播是一种作用于聚集和联合的工具，用于描述复合对象或集合对象对成员对象的依赖性并获得成员对象的属性的过程。它通过一种强制性的手段将成员对象的属性信息传播给复合对象。

复合对象的某些属性不需单独存储，可以从成员对象中提取或派生。成员对象的相关属性只能存储一次。这样，就可以保证数据的一致性，减少数据冗余。从成员对象中派生复合对象或集合对

象的某些属性值，其公共操作有“求和”、“集合和”、“最大”、“最小”、“平均值”和“加权平均值”等。例如，一个国家最大城市的人口数是这个国家所有城市人口数的最大值，一个省的面积是这个省所有县的面积之和，等等。

继承和传播在概念和使用上都是有差别的。这主要表现在：继承是用概括(“即是”关系)体系来定义的，服务于概括，而传播是用聚集(“成分”关系)或联合(“成员”关系)体系来定义的，作用于联合和聚集；继承是从上层到下层，应用于类，而传播是自下而上，直接作用于对象；继承包括属性和操作，而传播一般仅涉及属性；继承是一种信息隐含机制，只要说明子类与父类的关系，则父类的特征一般能自动传给它的子类，而传播是一种强制性工具，需要在复合对象中显式定义它的每个成员对象，并说明它需要传播哪些属性值。

3. 面向对象的空间数据模型

(1)基本含义。任何一种模型都无法反映现实世界的所有方面，对于复杂的事物和现象更是如此。因而不可能设计出一种通用的数据结构和数据模型来适应所有的情况，往往是在描述一类问题时，体现了优越性，而在描述另一类问题时，却是低效的。

为了有效地描述复杂的事物或现象，需要在更高层次上综合利用和管理多种数据结构和数据模型，并用面向对象的方法进行统一的抽象。这就是面向对象数据模型的含义，其具体实现就是面向对象的数据结构。

(2)复杂对象及其特点。面向对象的空间数据模型的核心是对复杂对象的模拟和操纵。

所谓复杂对象是指具有复杂结构和操作的对象。复杂对象可以由多种关系聚合抽象而成，或由不同类型的对象构成，或具有复杂的嵌套关系等。

例如，在GIS中的一个复杂地理实体(如学校)可能含有矢量数据、栅格数据、属性数据，甚至多媒体数据，而且可以认为是由其

他较简单的实体(如道路、教学楼、操场等)组成,因此,可以作为一个复杂对象。GIS 的地理实体所具有的矢量数据也可以认为是一个复杂对象,因为它包含了几何数据和属性数据,而几何数据又是由点、线、面等简单对象组成的。

因此,复杂对象的特点可归结为:

——一个复杂对象由多个成员对象构成,每个成员对象又可参与其他对象的构成;

——具有多种数据结构,如矢量、栅格、关系表等;

——一个复杂对象的不同部分可由不同的数据模型所支持,也就是说,可以分布于不同的数据库中。

(3)面向对象空间数据模型的特点。

——可充分利用现有数据模型的优点,面向对象的数据模型是一种基于抽象的模型,允许设计者在基本功能上选择最为适用的技术。如可以把矢量和栅格数据结构统一为一种高层次的实体结构,这种结构可以具有矢量结构和栅格结构的特点,但实际的操作仍然是矢量数据用矢量运算,栅格数据用栅格算法。

——具有可扩充性。由于对象是相对独立的,因此可以很自然和容易地增加新的对象,并且对不同类型的对象具有统一的管理机制。

——可以模拟和操纵复杂对象。传统的数据模型是面向简单对象的,无法直接模拟和操纵复杂实体,而面向对象的数据模型具备对复杂对象进行模拟和操纵的能力。

(4)GIS 中面向对象的数据模型。对于传统的数据模型,其数据结构是与之分离的。而对于面向对象的数据模型,其数据模型和数据结构是一致的,数据模型的具体实现就是数据结构。

在 GIS 的面向对象的数据结构中,通常可以把空间数据抽象为点、线、面三种简单的地物类型,作为三种简单对象:

点状地物 — 如塔、车站等。具有标识号、编码、定位点坐标等数据项,并且有显示、增加、删除、修改等操作。当点状地物是有向

地物时，需要两对定位点坐标。

线状地物——如道路、河流等。线状地物由一条或多条弧段组成，弧段还涉及两端的结点。具有标识号、编码、组成线状地物的弧段等数据项，具有显示、增加、删除、修改、计算长度等操作。

面状地物——如湖泊、街区等。由一条或多条弧段构成。除了具有标识号、编码、弧段串等数据项外，还常常含有面的内点坐标、面的面积，甚至面状地物的外接矩形的坐标等。除了具有增加、删除、修改、显示等操作外，还应具有计算面积的操作。

点、线、面这三种简单地物类型涉及到了孤立点、节点、结点、弧段等数据类型，并且还应该与注记有紧密的联系。

一个地理实体可以由这三种简单对象之一构成，复杂的地理实体可以由多种简单对象构成，甚至可以由其它复杂对象构成。每个地理实体都可以通过其标识号和其属性数据联系起来。若干个地理实体可以作为一个图层，若干个图层可以组成一个工作区。在GIS中可以开设多个工作区。

在GIS中建立面向对象的数据模型时，对象的确定还没有统一的标准，但是，对象的建立应符合人们对客观世界的理解，并且要完整地表达各种地理对象，及它们之间的相互关系。

4. 面向对象的数据库系统

(1)面向对象数据库系统的基本概念。关于"什么是面向对象数据库系统(OODBS－Object Oriented Data Base System)?" Francois Bancilho把OODBS定义为："一个面向对象的数据库系统应该满足两条准则：它应该是一个数据库管理系统，而且还是一个面向对象的系统。第一条准则是说它应该具备六个特征：永久性、外存管理、数据共享(并发)、数据可靠性(事务管理和恢复)、即席查询工具和模式修改。第二条准则是说它应具备八个特征：类/类型、封装性/数据抽象、继承性、多态性/滞后联编、计算完备性、对象标识、复杂对象和可扩充性。"

Rainer Unland在Francois Bancilho的基础上作了进一步的

补充，他指出："OODBS 除了须满足以上所定义的准则外，还必须满足非传统应用领域所提出的一些需求/特征，例如版本管理和可扩充的事务模型等等。"

Wan Kim 对 OODBS 的定义与他们的定义基本类似，他认为"一个数据库系统可以称为面向对象数据库系统，必须满足以下两个条件：支持核心的面向对象数据模型；借助扩充/修改传统数据库的语义，使之与核心的面向对象数据模型的语义一致，以支持传统数据库系统中所有的数据库特征。"并指出："面向对象数据库系统是一个面向对象数据库的永久、共享存贮器和管理程序；面向对象数据库是由面向对象数据模型定义的对象集合，该集合中的对象反映了面向对象程序设计所支持的对象语义。"

从 OODBS 具有面向对象特性的角度出发，OODBS 应该提供创建类的设施，用以组织对象、创建对象、把类组织成一个继承层次，使得子类能从超类中继承属性和方法，以及调用方法来访问特定的对象。

从 OODBS 是一个数据库系统的角度考虑，它必须提供当今关系数据库系统(RDB)提供的那些标准数据库设施，包括检索对象的非过程性查询设施、自动查询优化和处理、动态模式改变(改变类定义和继承结构)、存取方法(如 B+树索引、可扩充散列、排序等)的自动管理以提高查询处理性能、自动事务管理、并发控制、从系统故障中恢复、安全和授权。

(2)面向对象数据库系统的特征。1989 年 12 月，在第一届演绎、面向对象数据库国际会议上，以 Malcolm Atkinson ，Francois Bancilhon ，David Dewitt 等人为代表的一批 OODB 专家发表了著名的"面向对象的数据库系统宣言"，提出了 OODBS 应该具备的主要性质和特征。这些特征分为三组：必备的(被称作面向对象的数据库系统所必须满足的特性)、可选的(为了使系统更完善可添加的而非必备的特性)和开放的(设计人员可以选择的特性)。其中必备的特征包括：

复杂对象。对简单对象运用各种构造符即可得到复杂对象。

对象标识。在带对象标识的模型中，对象独立于它的值而存在，由此就有了两个对象等价的概念：两个对象可能相同（它们是同一个对象），也可能相等（它们有同样的值）。

封装性。封装性的概念来自两个方面，一是需要明确区分操作的说明和实现；二是模块化的需要。封装提供了一种逻辑数据独立性的形式，能够在不改变使用类型的程序的情况下改变类型的实现。让操作可见而将数据和操作的实现隐藏在对象中，就可以获得适当的封装。

类型和类。对应了两大类面向对象的系统：支持类概念的系统和支持类型概念的系统。类型和类具有相同的说明方式，类型的实例是静态说明的，而类的说明是动态产生的，它们都对应了抽象数据类型。

类和类型的层次结构。类和类型的层次结构可以导致类和类型的继承机制，在面向对象的系统中，至少存在有四种继承：替代继承、包含继承、限制继承和特化继承。

复载、过载和滞后联编。按照类型的不同重新定义每个类型的操作的实现叫做复载（Overriding），用同一个名字来表示不同的实现叫做过载（Overloading），它们都是由于多态性（Polymorphism）导致了同一方法名对应不同的对象而产生的调用不同的程序体的现象。为了实现这种功能，系统不能在编译时就把操作名联编到程序上，而要等到运行时才进行解释，这个推迟的转换叫做滞后联编（Late Binding）或动态联编（Dynamic Binding）。

计算完备性。就是能够用数据库系统的数据操纵语言（DML）表达任何可计算的功能。一般情况下，计算完备性可以通过与现有程序设计语言的合理连接来实现。

可扩充性。是指用户在数据库系统带有的一组预先定义类型的基础上所具有的可定义新类型的手段，并且系统定义的类型和

用户定义的类型在使用上没有区别。

持久性。持久性是指程序员保证他的数据在进程的执行中存在的能力。

辅存管理。辅存管理是数据库管理系统的一个经典特征。它通常有一组方法支持，包括索引管理、数据聚集、数据缓冲、存取路径选择及查询优化。

并发性。并发性是指数据库中的对象可被多个用户共享而不影响其一致性和完整性，并发性保证了同时在数据库上工作的多个用户之间的和谐共存。

恢复。恢复是指在硬件和软件发生故障时，数据库具有的能够使它返回到数据的某种相关状态的能力。

即席查询功能。就是数据库提供的一种高级的（即简明地表示重要的查询）、有效的（即查询的形式应该是进行过查询优化的）和与应用无关的（即它是独立于用户的特殊应用的）查询功能。

可选的特性有：多重继承性、类型检查和类型推理、分布、设计事务处理和版本等五个方面。

开放的特性是程序设计范例、表示系统、类型系统和一致性。

(3)面向对象数据库系统所具有的优势。与传统的数据库相比，OODB 在下列方面具有一定的优势：

缩小了语义差距。传统数据库设计往往是在问题空间采用某种语义模型（例如 E－R 模型），而在求解空间采用关系模型，于是就必须在这两个空间的表示之间作一个转换，这样往往会丢失语义。OODB 的优势在于在这两个空间中采用了相同/近似的模型，从而使它们之间的语义差距缩小了。

减轻了"阻抗失配"问题。传统数据库应用往往表现为把数据库语句嵌入某种具有计算完备性的程序设计语言中，由于数据库语言和程序设计语言的类型系统和计算模型往往不同，所以这种结合是不自然的，这个现象被称为"阻抗失配"。在 OODB 中，把需要程序设计语言编写的操作都封装在对象的内部，从本质上讲，

OODB 的问题求解过程只需要表现为一个消息表达式的集合。

适应非传统应用的需要。众所周知,OODB 研究的目的就是为了适应诸如 CAD、CAM、CASE、GIS 等非传统领域的需要。OODB 中,这种适应性主要表现在能够定义和操纵复杂对象,具备引用共享和并发共享机制以及灵活的事务模型(例如长事务模型、嵌套事务模型、切分事务模型),支持大量对象的存储和获取等等。

§3.4 空间数据库的设计、建立与维护

1. 数据管理概述

数据管理指的是对数据的分类、组织、编码、储存、检索和维护。数据管理在 30 多年来经历了三个阶段:人工管理阶段、文件系统阶段、数据库系统阶段。

(1)人工管理阶段(20 世纪 50 年代中期以前)。背景:计算机主要用于科学计算;从硬件看,外存只有磁带、卡片、纸带;从软件看,没有操作系统。

特点:数据不保存;没有软件系统对数据进行管理;基本上没有文件概念,数据的组织方式由程序员自行设计;一组数据对应一个程序,数据是面向应用的。

(2)文件系统阶段(20 世纪 50 年代后期到 60 年代中期)。背景:计算机除了科学计算外,还大量用于管理;从硬件看,外存有了磁盘、磁鼓等直接存取的存储设备;从软件看,操作系统已经有了专门的管理数据软件,即文件系统。

特点:计算机大量用于数据处理,数据需要长期保存;有软件进行数据管理;文件已经多样化;数据的存取基本上以记录为单位。数据冗余度大,数据和程序缺乏独立性。

(3)数据库系统阶段:(从 20 世纪 60 年代后期开始)。背景:计算机用于管理的规模更为庞大,共享的需要日益增强;有了大容量

的磁盘;联机实时处理要求增多,并开始考虑分布处理;软件价格上升,硬件价格下降。

特点:面向全组织的复杂的数据结构;数据冗余度小,易扩充;具有较高的数据和程序独立性;(·物理独立性:数据的存储结构(物理结构)改变时,数据的逻辑结构不变;·逻辑独立性:数据的总体逻辑结构改变时,数据的局部逻辑结构可以不变。)统一的数据控制功能;(·数据的安全性控制;·数据的完整性控制;·并发控制数据的最小存取单位是数据项。

2. 空间数据库的设计

数据库因不同的应用要求会有各种各样的组织形式。数据库设计就是根据不同的应用目的和用户要求,在一个给定的应用环境中,确定最优的数据模型、处理模式、存贮结构、存取方法,建立能反映现实世界的地理实体间信息之间的联系,满足用户要求,又能被一定的DBMS接受,同时能实现系统目标并有效地存取、管理数据的数据库。简言之,数据库设计就是把现实世界中一定范围内存在着的应用数据抽象成一个数据库的具体结构的过程。

空间数据库的设计是指在现有数据库管理系统的基础上建立空间数据库的整个过程。主要包括需求分析、结构设计和数据层设计三部分。

(1)需求分析。需求分析是整个空间数据库设计与建立的基础,主要进行以下工作:

——调查用户需求:了解用户特点和要求,取得设计者与用户对需求的一致看法。

——需求数据的收集和分析:包括信息需求(信息内容、特征、需要存储的数据)、信息加工处理要求(如响应时间)、完整性与安全性要求等。

——编制用户需求说明书:包括需求分析的目标、任务、具体需求说明、系统功能与性能、运行环境等,是需求分析的最终成果。

需求分析是一项技术性很强的工作,应该由有经验的专业技

术人员完成，同时用户的积极参与也是十分重要的。

(2)结构设计。指空间数据结构设计，结果是得到一个合理的空间数据模型，是空间数据库设计的关键。空间数据模型越能反映现实世界，在此基础上生成的应用系统就越能较好地满足用户对数据处理的要求。其主要任务是：

——概念设计。对需求分析阶段所收集的信息和数据进行分析、整理，确定地理实体、属性及它们之间的联系，将各用户的局部视图合并成一个总的全局视图，形成独立于计算机的反映用户观点的概念模式。概念模式与具体的 DBMS 无关，结构稳定，能较好地反映用户的信息需求。

表示概念模型最有力的工具是 E—R 模型，即实体—联系模型，包括实体、联系和属性三个基本成分。用它来描述现实地理世界，不必考虑信息的存储结构、存取路径及存取效率等与计算机有关的问题，比一般的数据模型更接近于现实地理世界，具有直观、自然、语义较丰富等特点，在空间数据库设计中得到了广泛应用。

例如在城市空间数据库系统设计中，我们将城市市区要素抽象为空间实体(街道边线、路段、街道、街区、节点等实体)、空间实体属性(节点实体属性——立交桥、警亭及所连通街道的性质等；边线实体属性——属于哪一路段、街道、街区及其长度等；街道路段和街道实体属性——走向、路面质量、宽度、等级、车道数、结构等；街区实体属性——面积、用地类型等)、空间实体关系。

近几年来，E—R 模型得到了扩充，增加了子类的概念，即增加了语义表达能力，使之能更好地模拟现实地理世界。

——逻辑设计。在概念设计的基础上，按照不同的转换规则将概念模型转换为具体 DBMS 支持的数据模型的过程，即导出具体 DBMS 可处理的空间数据库的逻辑结构(或外模式)，包括确定数据项、记录及记录间的联系、安全性、完整性和一致性约束等。导出的逻辑结构是否与概念模式一致，能否满足用户要求，还要对其功能和性能进行评价，并予以优化。

从 E—R 模型向关系模型转换的主要过程为：

①确定各实体的主关键字；

②确定并写出实体内部属性之间的数据关系表达式，即某一数据项决定另外的数据项；

③把经过消冗处理的数据关系表达式中的实体作为相应的主关键字；

④根据②、③形成新的关系。

⑤完成转换后，进行分析、评价和优化。

(3)物理设计。物理设计是指有效地将空间数据库的逻辑结构在物理存储器上实现，确定数据在介质上的物理存储结构，其结果是导出空间数据库的存储模式(内模式)。主要内容包括确定记录存储格式，选择文件存储结构，决定存取路径，分配存储空间。物理设计的好坏将对空间数据库的性能影响很大，一个好的物理存储结构必须满足两个条件：一是空间数据占有较小的存储空间；二是对数据库的操作具有尽可能高的处理速度。在完成物理设计后，要进行性能分析和测试。

数据的物理表示分两类：数值数据和字符数据。数值数据可用十进制或二进制形式表示。通常二进制形式所占用的存贮空间较少。字符数据可以用字符串的方式表示，有时也可利用代码值的存贮代替字符串的存储。为了节约存贮空间，常常采用数据压缩技术。

物理设计在很大程度上与选用的数据库管理系统有关。设计中应根据需要，选用系统所提供的功能。

(4)数据层设计。大多数 GIS 都将数据按逻辑类型分成不同的数据层进行组织。数据层是 GIS 中的一个重要概念。GIS 的数据可以按照空间数据的逻辑关系或专业属性分为各种逻辑数据层或专业数据层，原理上类似于图片的叠置。例如，地形图数据可分为地貌、水系、道路、植被、控制点、居民地等诸层分别存贮。将各层叠加起来就合成了地形图的数据。在进行空间分析、数据处理、图

形显示时，往往只需要若干相应图层的数据。

数据层的设计一般是按照数据的专业内容和类型进行的。数据的专业内容的类型通常是数据分层的主要依据，同时也要考虑数据之间的关系。如需考虑两类物体共享边界（道路与行政边界重合、河流与地块边界的重合）等，这些数据间的关系在数据分层设计时应体现出来。

不同类型的数据由于其应用功能相同，在分析和应用时往往会同时用到，因此在设计时应反映出这样的需求，即可将这些数据作为一层。例如，多边形的湖泊、水库，线状的河流、沟渠，点状的井、泉等，在 GIS 的运用中往往同时用到，因此，可作为一个数据层。

3. 空间数据库的建立与维护

(1) 空间数据库的建立。在完成空间数据库的设计之后，就可以建立空间数据库。建立空间数据库包括三项工作，即建立数据库结构、装入数据和试运行。

——建立空间数据库结构。利用 DBMS 提供的数据描述语言描述逻辑设计和物理设计的结果，得到概念模式和外模式，编写功能软件，经编译、运行后形成目标模式，建立起实际的空间数据库结构。

——数据装入。一般由编写的数据装入程序或 DBMS 提供的应用程序来完成。在装入数据之前要做许多准备工作，如对数据进行整理、分类、编码及格式转换（如专题数据库装入数据时，采用多关系异构数据库的模式转换、查询转换和数据转换）等。装入的数据要确保其准确性和一致性。最好是把数据装入和调试运行结合起来，先装入少量数据，待调试运行基本稳定了，再大批量装入数据。

——调试运行。装入数据后，要对空间数据库的实际应用程序进行运行，执行各功能模块的操作，对空间数据库系统的功能和性能进行全面测试，包括需要完成的各功能模块的功能、系统运行的

稳定性、系统的响应时间、系统的安全性与完整性等。经调试运行，若基本满足要求，则可投入实际运行。

由以上不难看出，建立一个实际的空间数据库是一项十分复杂的系统工程。

(2) 空间数据库的维护。建立一个空间数据库是一项耗费大量人力、物力和财力的工作，都希望能应用得好，生命周期长。而要做到这一点，就必须不断地对它进行维护，即进行调整、修改和扩充。空间数据库的重组织、重构造和系统的安全性与完整性控制等，就是重要的维护方法。

——空间数据库的重组织。指在不改变空间数据库原来的逻辑结构和物理结构的前提下，改变数据的存储位置，将数据予以重新组织和存放。因为一个空间数据库在长期的运行过程中，经常需要对数据记录进行插入、修改和删除操作，这就会降低存储效率，浪费存储空间，从而影响空间数据库系统的性能。所以，在空间数据库运行过程中，要定期地对数据库中的数据重新进行组织。DBMS 一般都提供了数据库重组的应用程序。由于空间数据库重组要占用系统资源，故重组工作不能频繁进行。

——空间数据库的重构造。指局部改变空间数据库的逻辑结构和物理结构。这是因为系统的应用环境和用户需求的改变，需要对原来的系统进行修正和扩充，有必要部分地改变原来空间数据库的逻辑结构和物理结构，从而满足新的需要。数据库重构通过改写其概念模式(逻辑模式)的内模式(存储模式)进行。具体地说，对于关系型空间数据库系统，通过重新定义或修改表结构，或定义视图来完成重构；对非关系型空间数据库系统，改写后的逻辑模式和存储模式需重新编译，形成新的目标模式，原有数据要重新装入。空间数据库的重构，对延长应用系统的使用寿命非常重要，但只能对其逻辑结构和物理结构进行局部修改和扩充，如果修改和扩充的内容太多，那就要考虑开发新的应用系统。

——空间数据库的完整性、安全性控制。空间数据库的完整

性，指数据的正确性、有效性和一致性，主要由后映象日志来完成，它是一个备份程序，当发生系统或介质故障时，利用它对数据库进行恢复。安全性指对数据的保护，主要通过权限授予、审计跟踪，以及数据的卸出和装入来实现。

§3.5 空间数据仓库

从信息系统诞生以来，人们生产了数以亿兆计的数据，它们以不同的格式存放在不同的地点，格式可有层次数据库、网状数据库、关系数据库、文件、电子表格等。这些数据分别由为支持组织日常运营而设计的各种信息系统来获取、存储和管理的。

传统的各种信息系统和联机事物处理系统被设计成能快速地、可靠地及有效地将数据送入数据库。在对数据库进行有意义的分析方面，这些系统并不太合适。存储在这些数据库中的数据是不一致且经常变化的，但分析者采用的方法需要工作处于稳定状态。由于数据库中明显地缺少历史数据，使得趋势分析几乎不可能。因此，由于数据来自面向应用、各种非集成化的数据源以及缺少历史数据，所以公司、部门实际存储和使用数据的能力受到了限制。尽管系统所收集的总积累数据量每隔两年就成倍增长，但这大量的数据并不能满足决策者的需要——因为他们是数据的富有者但却是信息的贫穷者。

建立数据仓库是决策支持系统(DSS)的最新演变，它基于数据库及网络技术的发展。数据仓库目前主要使用关系型数据库管理系统，是专为支持管理决策而设计的，并不是为了满足事物处理系统的需求。数据仓库提供了特殊的决策支持数据库，以此来管理从现有公司、部门的数据库及外部数据源到最终用户决策支持所要应用的信息流。数据仓库中存储了从业务系统和外部数据源中抽取的历史数据。这些业务及外部数据都是经过“净化”了的，消除了不一致性，并加以集成，从而创建了一个新的更适合于最终用户

进行业务分析的信息数据库。

1. 数据仓库

(1)数据仓库的概念。数据仓库是20世纪90年代发展起来的一种数据存储、管理、处理的技术。数据仓库有很多定义,如:"数据仓库是支持管理的、面向决策过程的、面向主题的、集成的、随时间变化的、持久的数据集合"(W. H. Inmon)。我们可以认为数据仓库就是网络数据库系统,具有如下特点:

——数据仓库的数据分布在用网络连接起来的不同数据库(可以是异构数据库)中;

——数据仓库的数据在来源、类型、格式、精度、编码等方面有很大差别;

——数据仓库使用元数据对数据进行求精、重构、管理;

——数据仓库可以对数据进行概括和聚集,为用户提供面向主题的数据;

——数据仓库能实现对不同数据源的同步更新;

(2)数据仓库的主要模块。

——数据求精:完成对数据的标准化、净化、过滤、匹配,对原始数据打上时间戳,确认数据源的数据质量。

——数据重构:对不同数据库中的各种数据进行集成、分割、概括、聚集、预测、推导、翻译、转换等,创建新数据。

——数据应用:提供各种数据分析工具、业务建模工具、决策支持工具、数据挖掘工具、数据查询与检索工具等。

2. 空间数据仓库

(1)空间数据仓库的概念。地理信息系统具有数据量巨大、包含多维空间属性和时间属性及其相关的特殊操作等特点。对数字地球的数据进行系统的整理、管理和深入的分析是未来地理信息系统研究的一个重要内容,是数字地球的基础研究及应用的重要前提之一。如何管理数字地球的海量数据和应付用户的频繁交互,如何给用户提供有用的、易于理解的空间地球信息。这需要通过空

间数据仓库技术、数据挖掘技术和数据融合技术等的帮助来实现。

空间数据仓库是数据仓库的一种特殊形式，是指支持管理和和决策过程的、面向主题的、集成的和随时间变化的、持久的地球数据的集合。

空间数据仓库根据一定的主题内容集成来自不同数据库中的数据，数据在结构上具有综合性；在时间上可以截取从瞬态到区段、直到全体等不同时间尺度上的信息；在属性上包括多个部门或专业应用系统的数据。空间数据仓库将数据的时间属性和空间属性紧密地结合起来，通过构建面向分析的多值空间数据模型，利用多维分析方法，从多个不同的角度进行分析比较，提取隐藏在数据中的信息，实现面向数据和面向模型的分析方法的统一。

空间数据仓库一般是非常庞大的数据库(VLDBS)，包容了数以百万或千万计的数据记录。由于数据是从不同的生产系统中搜集而来的，因而为进行空间分析建立了一个历史基础。为保持最新及准确的数据库，空间数据仓库要定期接收增加的新数据。定期加入新数据及删除不再需要的数据被称为数据的净化，这在刷新周期中进行。刷新周期必须快速、有效、且自动化，否则就丧失了数据仓库的根本价值。数据净化工具可将来自多处的数据合并到一个数据库中，自动搜集数据并检验、删除不需要的数据，从而维护数据库中的数据。

建立空间数据仓库的最大好处在其可用全新的、创造性的方式来关联数据。然而，复制数据库和空间数据仓库都增加了数据的冗余，这导致难以控制数据的完整性和准确性。要建立一个空间数据仓库是异常困难的。

(2)空间数据仓库的多维数据模型。空间数据仓库利用多维分类机制组织大量的空间数据，建立立方或超立方数据模型。维的数目由需求来定。空间查询一般分时间维、空间维、主题维(属性维)来组织数据。

通常按层次来组织维。分割不同的等级(精度)或层次，例如：

时间维：可划分为年、旬、月、周、日、时、分、秒等层次；

空间维：可划分为洲、国家、地区、省、市(县)、乡(镇)、村、户等层次；

主题维：可划分为类、亚类、种、亚种。

要实现空间数据仓库需要进行以下几项工作：

· 从不同数据源中，即从分布式的不同地点、不同部门的数据库中如何通过网络(Internet/Intranet)集成空间地球数据和元数据(Metadata)，如何解决它们在类型、格式、精度、编码等存在的巨大差别，如何解决分布式数据库的异构问题。

· 数据质量的精确和简练。

· 数据的概括和聚集，并非所有的数据都适合于空间数据仓库的查询分析要求，可以通过标准化、净化、过滤与匹配，为所选数据源中的原始数据打上时间戳进行求精，对来自分布式数据库中的数据进行集成与分割、概括与聚集、预测与推导、翻译与格式化、转换与再映像等重构工程，创建空间数据仓库所需要的数据子集。

· 元数据(Metadata)的创建和管理。元数据是对空间地球数据集的描述，是空间地球信息的标识、质量、结构、空间参照系、实体属性、数据变化(数据编码、数据格式、数据结构的变化)的历史等信息的代理，是用于数据描述和数据组织的数据。

· 数据源(分布在任何地点的数据库)的同步更新问题，必须对已存储数据进行转移或再映像到原始数据中，以便对最新衍生、创建或转换的数据不断进行更新。

3. 空间数据的挖掘和融合

(1)空间数据挖掘。数据挖掘是近年来出现的一种信息分析工具。数据挖掘能自动地找出数据仓库中的模式及关系。数据挖掘代表了决策支持解决方案的下一发展趋势。它利用了先进的统计技术及设备，力图在大型数据库(包括在因特网上的数据库)中找出各项事实。与要求用户给出公式并检验某一假设的查询工具不同，数据挖掘利用了内嵌的分析算法，可在用于预测将来活动的数

据中自动生成有关模式和异常事件的假设。除了采用模式识别及分析的新技术外，数据挖掘还采用了许多传统的技术，如关系模型DBMS、OLAP、大批量并行处理(MPP)、对称多重处理(SMP)及统计学。

对于许多最终用户目前尚未想到的问题，数据挖掘工具也能给出答案。具体做法是通过描绘数据内在模式的特点，从这些模式中开发出一个假设并运用假设来预测未来的行为。数据挖掘告诉用户为什么会发生这些事件。当你使用数据挖掘工具时，你根本不必详细地描绘出你所要问的问题，相反，你只要说，"这些是数据，请告诉我其中会有哪些有意义的模式"。

数据挖掘也只对它所能存取的数据有效，数据仓库为有效的数据挖掘扫清了障碍。数据挖掘提供了自底向上的，发现式驱动的数据分析。数据仓库需要对源于用户的理论反复测试，而数据挖掘根本不需要任何假设，甚至可以根据发现的模式来识别事实和结论。

美国正启动一个空间信息处理项目(Earth Overview System，EOS)，能细粒度地监测地面每平方公里上发生的情况，该项目对巩固美国在全球的竞争优势具有重要的作用。该项目的主要组成部分之一就是空间数据的联机分析与挖掘技术的研究。

空间数据挖掘的研究内容：

——支持数字地球的空间数据联机分析处理技术，根据数字地球的联机分析要求，研究如何高效地进行空间数据的联机分析。主要包括以下几个方面：

- 空间 cube 的高效计算方法；
- 运用空间数据 cube 的联机分析处理技术。

——如何运用数据仓库技术进行以数字地球为对象的深层次数据分析和知识获取，即所谓的空间数据挖掘。由于空间数据的特殊性，在进行空间数据挖掘时，采用可视化的技术手段是非常必要的。以下方面是需要进行研究的：

• 空间数据可视化分析；
• 空间数据关联分析；
• 空间数据聚集分析；
• 空间数据分类算法；
• 空间数据联机集成挖掘。

—— 空间数据仓库和空间数据联机分析的 Internet 发布，研究把空间数据仓库中的信息和空间数据联机分析、空间数据挖掘的结果通过 Internet 进行发布。主要包括以下方面：

• Web 空间数据仓库的模型与机制；
• 基于 Web 的联机空间数据分析处理。

(2)空间数据融合。信息融合也是信息技术的发展趋势之一，数字化、信息化和网络化为信息融合提供了技术基础。数字地球为地球空间信息和各种属性信息的融合提供了框架。

—— 数字地球信息融合框架的时空模型。空间经度、纬度、高度以及时间构成框架的四维坐标。地球形状和表面是不规则的。采用标准椭球体作为参照，定义经纬度，以及相应的大地水准面来定义高度。另外也可以采用地心坐标系统，由 X，Y，Z 以及时间 T 构成四维坐标，空间时间因子是基础的定位因子，确定其他因子的时空分布。GPS 可提供定位数据。在地球坐标系统之间可实现相互转换。

空间信息融合框架是围绕信息基础设施的相关技术而建立的，主要围绕 Internet 以及 NGI(下一代 Internet)而融合多种信息资源，包括各种计算设备、日用电子产品，GPS 接收机。这些设备可以通过 Internet 访问数字地球中的信息，或为数字地球输入信息。

——空间信息融合框架所需的关键技术。

宽带网：由于数字地球将由成千上万个数据库组成，需要成千上万个不同组织来维护，这意味着组成数字地球的服务器需要由高速网连接。NGI 可提供 1000Gbps 的带宽。

海量存储:图象数据库将存储每天产生的1000G字节的地球观测数据。

互操作:如何实现成千上万的数据库的联合,实现自然、环境、社会、经济和人口信息的互操作将是巨大的挑战。

第四章　空间数据的数学基础

§4.1　空间数据的地理参照系

目前，世界各国、各地区已建立了各种规模和类型的GIS，这些系统为经济、国防等各个领域的科学决策提供了依据，发挥了重要的作用。但是不论每个应用型的GIS的服务目的是什么，每个GIS自身的数据必须是在统一的地理参照系下的数据，也就是说要有统一的坐标系和高程系。

1. 地球的形状

地球近似球体，其表面高低不平，极其复杂。

假想将静止的平均海水面延伸到大陆内部，可以形成一个连续不断的、与地球比较接近的形体。把该形体视为地球的形体，其表面就称之为大地水准面。

但是，由于地球内部物质分布不均匀和地面高低起伏不平，使各处的重力方向发生局部变异，处处与重力方向垂直的大地水准面显然不可能是一个十分规则的表面，且不能用简单的数学公式来表达，因此，大地水准面不能作为测量成果的计算面。

为了测量成果计算的需要，选用一个同大地体相近的、可以用数学方法来表达的旋转椭球来代替地球，且这个旋转椭球是由一个椭圆绕其短轴旋转而成的。

凡是与局部地区（一个或几个国家）的大地水准面符合得最好的旋转椭球，称为参考椭球。经过长期的观测、分析和计算，世界上许多学者和机构算出了参考椭球的长短半径的数值。

我国1952年前采用的是海福特椭球，1953年起改用克拉索夫斯基椭球。1978年后开始采用1975年国际椭球，并以此建立了我国新的、独立的大地坐标系。

2. 坐标系

坐标系——确定地面点或空间目标位置所采用的参考系。与测量相关的主要有地理坐标系和平面坐标系。

(1)地理坐标系。地理坐标系——用经纬度来表示地面点的位置。地面上任一点M的位置可由经度λ和纬度φ来决定，记为M(λ,Φ)。

经纬度具有深刻的地理意义，它标示物体在地面上的位置，显示其地理方位(经线与南北相应，纬线与东西相应)，表示时差，此外，经纬线还标示许多地理现象所处的地理带，如气候、土壤等部门都要利用经纬度来推断地理规律。

经纬度的测定方法主要有两种，即天文测量和大地测量。以大地水准面和铅垂线为依据，用天文测量的方法，可获得地面点的天文经纬度。测有天文经纬度坐标（λ，Φ）的地面点，称为天文点。

以旋转椭球和法线为基准，用大地测量的方法，根据大地原点和大地基准数据，由大地控制网逐点推算各控制点的坐标(L,B)，称为大地经纬度。

解放前，我国实际上没有统一的大地坐标系。解放初期从原苏联1942年坐标系经联测和平差计算引伸到我国，建立了1954年北京坐标系。该坐标系的椭球面与我国大地水准面不能很好地符合，产生的误差较大，不能满足我国空间技术、国防尖端技术、经济建设的要求。

我国在积累了30年测绘资料的基础上，通过全国天文大地网整体平差建立了我国的大地坐标系。该坐标系采用1975年国际椭球参数，国家大地原点设在陕西省。该系统坐标统一、精度优良，可直接满足1:5000甚至更大比例尺测图的需要。

(2)平面坐标系。将椭球面上的点通过投影的方法投影到平面上时，通常使用平面坐标系。平面坐标系分为平面极坐标系和平面直角坐标系。

平面极坐标系采用极坐标法，即用某点至极点的距离和方向来表示该点的位置的方法，来表示地面点的坐标。主要用于地图投影理论的研究。

平面直角坐标采用直角坐标(笛卡尔坐标)来确定地面点的平面位置。可以通过投影将地理坐标转换成平面坐标。

3. 高程系

高程——由高程基准面起算的地面点的高度。

而高程基准面是根据多年观测的平均海水面来确定的。也就是说，高程(也称海拔高程、绝对高程)是指地面点至平均海水平的垂直高度。

地面点之间的高程差，称为相对高程，简称高差。

由于不同地点的验潮站所得的平均海平面之间存在着差异，所以，选用不同的基准面就有不同的高程系统。

一个国家一般只能采用一个平均海水面作为统一的高程基准面。我国的高程基准原来采用“1956 年黄海高程系”，由于观测数据的积累，黄海平均海水面发生了微小的变化，因此启用了新的高程系，即“1985 年国家高程基准”。

在采用新的高程基准后，对已有地图的等高线高程的影响可忽略不计。

§4.2 地图投影

1. 地图投影的实质

不规则的地球表面可以用地球椭球面来替代。由于地球椭球面是不可展曲面，而地图是一个平面，因而将地球椭球面上的点映射到平面上来的方法，称为地图投影。

对于较小区域范围,可以视地表为平面,可以认为投影没有变形。但对于大区域范围,甚至是半球、全球,这种投影方法就不太适合了。这时,可以考虑另外的投影方法,例如,可以假设地球按比例尺缩小成一个透明的地球仪那样的球体,在其球心、球面或球外安放一个发光点,将地球仪上经纬线(连同控制点及地形、地物图形)投影到球外的一个平面上,即成为地图。图 4—1 是将地球表面投影在圆柱面上的透视投影示意图。

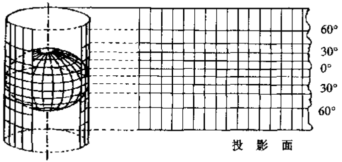


图 4—1 透视投影示意图

实际上这种直观的透视投影方法亦有很大的局限性,例如,往往不能将全球投影下来,多数情况下不可能用这种几何作图的方法来实现。科学的投影方法是建立地球椭球面上的经纬线网与平面上相应的经纬线网相对应的基础上的,其实质就是建立地球椭球面上点的坐标(λ,φ)与平面上对应的坐标(x,y)之间的函数关系,用数学表达式表示为:

$$x=f_1(\lambda,\varphi)$$
$$y=f_2(\lambda,\varphi) \qquad (4—1)$$

这是地图投影的一般方程式。当给定不同的具体条件时,就可得到不同种类的投影公式。

2. 投影变形

由于要将地球椭球面展开成平面,且不能有断裂,那么图形必

将在某些地方被拉伸，某些地方被压缩，因而投影变形是不可避免的。

投影变形通常包括三种，即长度变形、角度变形和面积变形。

长度变形(ν_μ)是长度比与1之差值，即

$$\nu_\mu=\mu-1 \quad (4-2)$$

而长度比(μ)则是指地面上微分线段投影后长度ds'与其固有长度ds之比。即

$$\mu=\frac{ds'}{ds}$$

长度比是一个变量，不仅随点位不同而变化，而且在同一点上随方向不同也有差异。

角度变形是指实际地面上的角度(α)和投影后角度(α')的差值，即

$$\alpha-\alpha' \quad (4-3)$$

角度变形可以在许多地图中均可清晰地看到。本来经纬线在实地上是成直角相交的，但经过投影之后，很多情况下经纬线变成了非直角相交的图形。

面积变形(v_p)是指面积比P与1之差，即

$$v_p=P-1 \quad (4-4)$$

式中P是面积比，是地球表面上微分面积投影后的大小dF'与其固有面积dF之比，即P=dF'/dF

面积比也是一个变量，它随点位不同而变化。因此，面积变形亦在许多投影中经常出现。

3. **投影的分类**

(1)按投影变形性质分类。根据地图投影中可能引入的变形的性质，可以分为等角投影、等面积投影和任意投影三种。

等角投影保证了投影后任意点的由任意两条微分线段构成的角度不产生变形，这种投影可以使得投影前后的形状保持不变，因而也称之为正形投影。

等面积投影保证了投影前后面积保持不变，对微分面积如此，对整个区域的较大面积亦如此。

任意投影在投影后既不保持角度不变，又不保持面积不变，它同时存在着长度、角度和面积的变形。在任意投影中，如果存在某一方向上长度不变时，我们将之称为等距离投影。

等角与等面积是相互抵触的，也就是说等角是以牺牲等面积为代价的；同样等面积也是以牺牲等角为前提的；任意投影虽然存在着各种变形，但各种变形比较均衡。

(2)根据投影面及其位置分类。在地图投影中，我们首先将不可展的地球椭球面投影到一个可展曲面上，然后将该曲面展开成为一个平面，得到我们所需要的投影。通常采用的这个可展曲面有圆锥面、圆柱面、平面(曲率为零的曲面)，相应地可以得到圆锥投

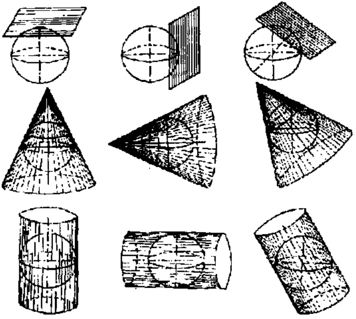


图 4-2　根据投影面及其位置分类的地图投影

影、圆柱投影、方位投影。同时我们还可以由投影面与地球轴向的相对位置区分为正轴投影(投影面的中心轴与地轴重合)、斜轴投影(投影面的中心轴与地轴斜向相交)、横轴投影(投影面中心轴与地轴相互垂直)。

各种投影都具有一定的局限性。一般地说,距投影面越近,变形就越小。为了控制投影的变形分布,我们可以调整投影面与椭球体面的相交位置,根据这个相交位置,又可以进一步得到各种投影相应的切投影(投影面与椭球体相切)和割投影(投影面与椭球体相割)。对这一体系的分类可以用图 4-2 给出。

此外,尚有的方位、伪圆锥、伪圆柱等许多类型的投影,此处从略。

4. GIS 中的地图投影

GIS 所存贮记录、管理分析、显示应用的内容是地理信息,而地理信息的描述必须要有指定的地理参照系,且地理位置应以地理坐标或平面坐标的方式表示出来。

地图不仅是 GIS 的重要数据源,而且是表示地理信息的最佳媒介。也就是说,在 GIS 中,地理信息基本上都是以地图的方式显示给用户的,用户也是在地图上进行空间信息的查询的,GIS 空间分析的结果也是以地图的形式显示出来的,GIS 输出的成果中大部分也是地图等等。

由于 GIS 大多是以地图的方式来显示地理信息的,而地图是平面,地理信息则是在地球椭球面上的,因此,地图投影在 GIS 中是不可缺少的。

例如,当 GIS 的地理数据库中的地理数据是以地理坐标(即经纬度)来存贮时,对于输入的以地图为数据源的空间数据必须通过投影变换来转换成地理坐标,然后才能装入到 GIS 的地理数据库中,而当需要显示或输出地图时,则必须将地理数据库中的地理坐标表示的空间数据通过投影变换成指定投影的平面坐标。

在 GIS 中,地理数据的显示往往可以根据用户的需要,指定

各种投影。但当所显示的地图与国家基本地图系列的比例尺一致时，往往采用与国家基本系列地图所用的投影。我国常用的地图投影的情况为：

(1)我国基本比例尺地形图(1:100 万、1:50 万、1:25 万、1:10 万、1:5 万、1:2.5 万、1:1 万、1:5000)除 1:100 万外均采用高斯—克吕格投影为地理基础；

(2)我国 1:100 万地形图采用了 Lambert 投影，其分幅原则与国际地理学会规定的全球统一使用的国际百万分之一地图投影保持一致；

(3)我国大部分省区图以及大多数这一比例尺的地图也多采用 Lambert 投影和属于同一投影系统的 Albers 投影(正轴等面积割圆锥投影)；

(4)Lambert 投影中，地球表面上两点间的最短距离(即大圆航线)表现为近于直线，这有利于地理信息系统中和空间分析量度的正确实施。

因此，下面把高斯—克吕格投影和正等角圆锥投影作一简要介绍。

5. 高斯—克吕格投影

高斯—克吕格投影是一种横轴等角切椭圆柱投影。它是将一椭圆柱横切于地球椭球体上，该椭圆柱面与椭球体表面的切线为一经线，投影中将其称为中央经线，然后根据一定的约束条件即投影条件，将中央经线两侧规定范围内的点投影到椭圆柱面上，从而得到点的高斯投影(图 4—3)。

高斯投影的条件为：

——中央经线和地球赤道投影成为直线且为投影的对称轴；

——等角投影；

——中央经线上没有长度变形。

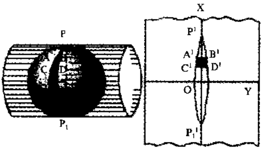


图 4—3 高斯投影及其图形

根据高斯投影的条件推导出的高斯—克吕格投影的计算公式为：

$$\begin{cases} X=S+\frac{\lambda^2N}{2}\sin\varphi\cos\varphi+\frac{\lambda^2N}{24}\sin\varphi\cos^3\varphi(5-tg^2\varphi+9\eta^2+4\eta^4)+\cdots\cdots \\ Y=\lambda N\cos\varphi+\frac{\lambda^3N}{6}\cos^3\varphi(1-tg^2\varphi+\eta^2)+\frac{\lambda^5N}{120}\cos^5\varphi(5-18tg^2\varphi \\ \quad +tg^4\varphi)+\cdots\cdots \end{cases} \tag{4—5}$$

式中：X、Y 为点的平面直角坐标系的纵、横坐标；

φ、λ 为点的地理坐标，以弧度计，λ 从中央经线起算；

S 为由赤道至纬度 φ 处的子午线弧长；

N 为纬度 φ 处的卯酉圈曲率半径；

其中 η 为地球的第二偏心率，a、b 则分别为地球椭球体的长短半轴。

$$\eta=e'\cos2\varphi$$
$$e'^2=(a^2-b^2)/b^2$$

高斯投影由于是等角投影，故没有角度变形，其沿任意方向的长度比都相等，其面积变形是长度的两倍。对高斯—克吕格投影长度变形的研究可以依下述长度比表达式进行：

$$\mu=1+\frac{1}{2}\cos^2\varphi(1+\eta^2)\lambda^2+\frac{1}{6}\cos^4\varphi(2-tg^2\varphi)\lambda^4-\frac{1}{8}\cos^4\varphi\lambda^4+\cdots\cdots$$

由该长度比公式可以分析出高斯投影变形具有以下特点：

· 中央经线上无变形；

· 同一条纬线上，离中央经线越远，变形越大；

· 同一条经线上，纬度越高，变形越大；

· 等变形线为平行于中央经线的直线。

由此可见，高斯投影的最大变形处为各投影带在赤道边缘处，为了控制变形，我国地形图采用分带方法，即将地球按一定间隔的经差(6°或 3°)划分为若干相互不重叠的投影带，各带分别投影。1：2.5 万至 1:50 万的地形图均采用 6°分带方案，即从格林尼治零度经线起算，每 6°为一个投影带，全球共分为 60 个投影带。我国领土位于东经 72°到 136°之间，共包括 11 个投影带(13 带～22 带)。1：1 万及更大比例尺地形图采用 3°分带方案，全球共分为 120 个投影带。图 3—4 给出了高斯投影的 6°带和 3°带分带方案。

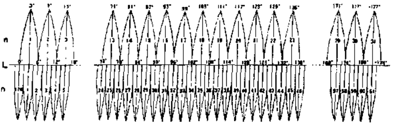


图 4—4　高斯投影分带方案

为了制作地图和使用地图的方便，通常在地图上都绘有一种或两种坐标网，即经纬线网和方里网。

经纬线网——即指由经线和纬线所构成的坐标网，又称地理坐标网。

在 1:1 万～1:20 万比例尺的地形图上，经纬线只以图廓线的形式直接表现出来，并在图角处注出相应度数。为了在用图时加密成网，在内外图廓间还绘有加密经纬网的加密分划短线(图式中称“分度带”)，必要时与对应短线相连就可以构成加密的经纬线网。

1:25 万地形图上，除内图廓上绘有经纬网的加密分划外，图内还有加密用的十字线。

我国的 1:50 万～1:100 万地形图，在图面上直接绘出经纬线网，内图廓上也有供加密经纬线网的加密分划短线。

方里网——是由平行于投影坐标轴的两组平行线所构成的方格网。因为是每隔整公里绘出坐标纵线和坐标横线，所以称之为方里网。由于方里线同时又是平行于直角坐标轴的坐标网线，故又称直角坐标网。

直角坐标网的坐标系以中央经线投影后的直线为 X 轴，以赤道投影后的直线为 Y 轴，它们的交点为坐标原点。这样，坐标系中就出现了四个象限。纵坐标从赤道算起向北为正、向南为负；横坐标从中央经线算起，向东为正、向西为负。

我国位于北半球，全部 X 值都是正值。在每个投影带中则有一半的 Y 坐标值为负。为了避免 Y 坐标出现负值，规定纵坐标轴向西平移 500km(半个投影带的最大宽度不超过 500km)。这样，全部坐标值都表现为正值了。

6. 正等角圆锥投影

这种投影，是假想圆锥轴和地球椭球体旋转轴重合并套在椭球体上，圆锥面与地球椭球面相割，将经纬网投影于圆锥面上展开而成。其经线表现为辐射的直线束，纬线投影成同心圆弧(图 4—5)。

圆锥面与椭球面相割的两条纬线圈，称之为标准纬线(φ_1，φ_2)。采用双标准纬线的相割比采用单标准纬线的相切，其投影变形小而均匀。

投影变形的分布规律是：

——角度没有变形，即投影前后对应的微分面保持图形相似，故亦可称为正形投影；

——等变形线和纬线一致，同一条纬线上的变形处处相等；

——两条标准纬线上没有任何变形；

——在同一经线上，两标准纬线外侧为正变形（长度比大于1），而两标准纬线之间为负变形（长度比小于1），因此，变形比较均匀，绝对值也较小；

——同一纬线上等经差的线段长度相等，两条纬线间的经线线段长度处处相等。

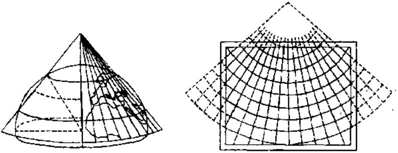


图4—5　正等角割圆锥投影及其经纬线图形

§4.3　地图投影变换

当系统所使用的数据是来之不同地图投影的图幅时，需要将一种投影的数字化数据转换成所需投影的数字化数据，这就需要进行地图投影变换。

地图投影变换的实质是建立两平面场之间点的一一对应关系。假定原图点的坐标为x，y（称为旧坐标），新图点的坐标为X，Y（称为新坐标），则由旧坐标变换为新坐标的基本方程式为：

$$X=f_1(x,y)$$
$$Y=f_2(x,y) \qquad (4-6)$$

实现由一种地图投影点的坐标变换为另一种地图投影点的坐标就是要找出上述关系式，其方法通常分为三类：

1. 解析变换法

这类方法是找出两投影间坐标变换的解析计算公式。由于所

采用的计算方法不同又可分为反解变换法和正解变换法。

反解变换法(又称间接变换法)。这是一种中间过渡的方法,即先解出原地图投影点的地理坐标 φ,λ,然后将其代入新图的投影公式中求得其坐标。即

$$\boxed{x,y} \longrightarrow \boxed{\varphi,\ \lambda} \longrightarrow \boxed{X,Y}$$

正解变换法(又称直接变换法)。这种方法不需要反解出原地图投影点的地理坐标,而是直接求出两种投影点的直角坐标关系式。即

$$\boxed{x,y} \longrightarrow \boxed{X,Y}$$

2. 数值变换法

如果原投影点的坐标解析式不知道,或不易求出两投影之间坐标的直接关系,可以采用多项式逼近的方法,即用数值变换法来建立两投影间的变换关系式。例如,可采用二元三次多项式进行变换。二元三次多项式为:

$$\begin{cases} X=a_{00}+a_{10}x+a_{01}y+a_{20}x^2+a_{11}xy+a_{02}y^2+a_{30}x^3+a_{21}x^2y \\ \quad +a_{12}xy^2+a_{03}y^3 \\ Y=b_{00}+b_{10}x+b_{01}y+b_{20}x^2+b_{11}xy+b_{02}y^2+b_{30}x^3+b_{21}x^2y \\ \quad +b_{12}xy^2+b_{03}y^3 \end{cases} \tag{4—7}$$

通过选择 10 个以上的两种投影之间的共同点,并组成最小二乘法的条件式,即:

$$\sum_{i=1}^{n}(X_i-X'_i)^2=\min$$

$$\sum_{i=1}^{n}(Y_i-Y'_i)^2=\min \tag{4—8}$$

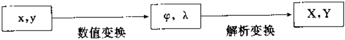
其中:n 为点数,X_i,Y_i 为新投影的实际变换值,X_i',Y_i' 为新投影的理论值。

根据求极值原理,可得到两组线性方程,即可求得各系数的

值。

3. 数值解析变换法

当已知新投影的公式，但不知原投影的公式时，可先通过数值变换求出原投影点的地理坐标 φ,λ，然后代入新投影公式中，求出新投影点的坐标。即



第五章　空间数据的输入

空间数据输入的任务是将现有的地图、外业观测成果、航空像片、遥感图像、文本资料等转换成 GIS 可以处理与接收的数字形式，通常要经过验证、修改、编辑等处理。

不同数据的输入需要用到不同的设备。例如，对于文本数据通常用交互的方式通过键盘录入，也可用扫描仪扫描后用字符识别软件自动录入；对于矢量地图数据，可用平板数字化仪，采用手扶跟踪的方法输入，也可用扫描仪扫描成图像后，用栅格数据矢量化的方法自动追踪输入等等。

GIS 软件的这一部分还应具有数据转换装载的功能，即能把其他 GIS 或专题数据库中的数据通过转换装载到当前的 GIS 系统中。

这一部分 GIS 软件的数据处理工作主要是几何纠正、图形和文本数据的编辑、图幅的拼接、拓扑关系的生成等，即完成 GIS 的空间数据在装入 GIS 的地理数据库前的各种工作。

§5.1　GIS 属性数据的编码

1. GIS 属性数据的含义

GIS 属性数据是描述空间数据的属性特征的数据。例如，道路可以数字化为栅格表示的一组连续的像元或矢量表示的线，而道路的属性数据是指道路的宽度、等级、表面类型、建筑方法、建筑日期、特殊的交通规则、车流量等等。这些数据可以存储在数据库中，通过唯一的标识符与相应的几何数据联系起来。

当属性数据的数据量较大时，通常与几何数据分开输入。首先将属性数据输入到一个文件，经编辑、检查无误后存入数据库。

在属性数据中，有一部分是与几何数据的表示密切有关的，例如，道路的等级、类型等，决定着道路符号的形状、色彩、尺寸等。在GIS中，通常把这部分属性数据用编码的形式表示，并与几何数据一起管理起来。编码的过程是将信息转换成数据的过程，前提是首先要对需表示的信息进行分类分级。

2. 属性数据的分类分级

(1)分类的基本原则。分类是将具有共同属性或特征的事物或现象归并在一起，而把不同属性或特征的事物或现象分开的过程。

分类是人类思维所固有的一种活动，是认识事物的一种方法。分类的基本原则是：

——科学性：选择事物或现象最稳定的属性和特征作为分类的依据。

——系统性：应形成一个分类体系，低级的类应能归并到高级的类中。

——可扩性：应能容纳新增加的事物和现象，而不致于打乱已建立的分类系统。

——实用性：应考虑对信息分类所依据的属性或特征的获取方式和获取能力。

——兼容性：应与有关的标准协调一致。

(2)分类的基本方法。

——线分类法：线分类法又称层级分类法。它是将初始的分类对象按所选定的若干个属性或特征依次分成若干个层级目录，并编排成一个有层次的、逐级展开的分类体系。其中同层级类目之间存在并列关系，不同层级类目之间存在隶属关系，同层类目互不重复、互不交叉。

线分类法的优点是容量较大，层次性好，使用方便；缺点是分类结构一经确定，不易改动，当分类层次较多时，代码位数较长。

——面分类法，面分类法是将给定的分类对象按选定的若干个属性或特征分成彼此互不依赖、互不相干的若干方面（简称面），每个面中又可分成许多彼此独立的若干个类目。使用时，可根据需要将这些面中的类目组合在一起，形成复合类目。

面分类法的优点是具有较大的弹性，一个面内类目的改变，不会影响其它面，且适应性强，易于添加和修改类目；缺点是不能充分利用容量。

(3)分级的基本原则。分级是对事物或现象的数量或特征进行等级的划分，主要包括确定分级数和分级界线。

——确定分级数的基本原则

• 分级数应符合数值估计精度的要求。分级数多，数值估计的精度就高。

• 分级数应顾及可视化的效果。等级的划分在GIS中要以图形的方式表示出来，根据人对符号等级的感受，分级数应在4～7级。

• 分级数应符合数据的分布特征。对于呈明显聚群分布的数据，应以数据的聚群数作为分级数。

• 在满足精度的前提下，应尽可能选择较少的分级数。

——确定分级界线的基本原则

• 保持数据的分布特征。使级内差异尽可能小，各级代表值之间的差异应尽可能大。

• 在任何一个等级内都必须有数据，任何数据都必须落在某一个等级内。

• 尽可能采用有规则变化的分级界线。

• 分级界线应当凑整。

(4)分级的基本方法。在分级时大多采用数学方法，如数列分级、最优分割分级等。对于有统一标准的分级方法时，应采用标准的分级方法，如按人口数把城市分为特大城市、大城市、中等城市、小城市等。也可以定性地分级，如国家、省、市、县、镇等。

3. 属性数据的编码

属性数据的编码是指确定属性数据的代码的方法和过程。代码是一个或一组有序的易于被计算机或人识别与处理的符号，是计算机鉴别和查找信息的主要依据和手段。编码的直接产物就是代码，而分类分级则是编码的基础。

(1)代码的功能。

——鉴别：代码代表对象的名称，是鉴别对象的唯一标识。

——分类：当按对象的属性分类，并分别赋予不同的类别代码时，代码又可作为区分分类对象类别的标识。

——排序：当按对象产生的时间、所占的空间或其他方面的顺序关系排列，并分别赋予不同的代码时，代码又可作为区别对象排序的标识。

(2)编码的基本原则。

——唯一性：一个代码只唯一地表示一类对象。

——合理性：代码结构要与分类体系相适应。

——可扩性：必须留有足够的备用代码，以适应扩充的需要。

——简单性：结构应尽量简单，长度应尽量短。

——适用性：代码应尽可能反映对象的特点，以助记忆。

——规范性：代码的结构、类型、编写格式必须统一。

(3)代码的类型。代码的类型是指代码符号的表示形式，有数字型、字母型、数字和字母混合型三类。

数字型代码——用一个或若干个阿拉伯数字表示对象的代码。特点是结构简单、使用方便、易于排序，但对对象的特征描述不直观。

字母型代码——用一个或若干个字母表示对象的代码。特点是比同样位数的数字型代码容量大，还可提供便于识别的信息，易于记忆，但比同样位数的数字型代码占用更多的计算机空间。

数字、字母混合型代码——是由数字、字母、专用符组成的代码。兼有数字型和字母型的优点，结构严密，直观性好，但组成形式复杂，处理麻烦。

(4)GIS 中代码的种类。GIS 中的代码可以分为两种,一种是分类码,另一种是标识码。

分类码——是根据地理信息分类体系设计出的各专业信息的分类代码,用以标识不同类别的数据,根据它可以从数据中查询出所需类别的全部数据。

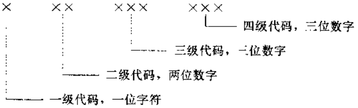
标识码(亦称识别码)——是在分类码的基础上,对每类数据设计出其全部或主要实体的识别代码,用以对某一类数据中的某个实体,如一个居民地、一条河流、一条道路等进行个体查询检索,从而弥补分类码不能进行个体分离的缺陷。标识码是联系实体的几何信息和属性信息的关键字。

(5)编码方法举例。

——行政区划代码(GB—2260—91)。这是一种识别码,用 6 位数字代码按层次分别表示省(自治区、直辖市)、地区(市、州、盟)、县(区、市、旗)的名称。其第一、二位表示省(自治区、直辖市);第三、四位表示省直辖市(地区、州、盟),其中 01～20,51～70 表示省直辖市,21～50 表示地区、州、盟;第五、六位表示县(市辖市、地辖市、县级市、旗),其中 01～18 表示市辖区或地辖市,21～80 表示县、旗,81～99 表示县级市。

例如:郑州市的代码为 410100

——加拿大数字地形要素分类编码系统。这是一种分类码,且是一种数字字母混合型代码。采用树型结构将地形要素分为四级,其代码结构为:



§5.2 空间数据的采集

GIS 的核心是地理数据库，所以，建立 GIS 的第一步就是将空间实体的几何数据和属性数据输入到地理数据库中，这就是 GIS 的数据采集。

GIS 需要输入两方面的数据，即几何数据与属性数据。为此需进行三方面的工作，即

(1)几何数据的采集；

(2)属性数据的采集；

(3)几何数据与属性数据的连接。

1. 几何数据的采集

在 GIS 的几何数据采集中，如果几何数据已存在于其他的 GIS 或专题数据库中，那么只要经过转换装载即可；对于由测量仪器获取的几何数据，只要把测量仪器的数据传输进入数据库即可，测量仪器如何获取数据的方法和过程通常是与 GIS 无关的。

对于栅格数据的获取，GIS 主要涉及使用扫描仪等设备对图件的扫描数字化，这部分的功能也较简单。因为通过扫描获取的数据是标准格式的图像文件，大多可直接进入 GIS 的地理数据库。

从遥感影像上直接提取专题信息，需要使用几何纠正、光谱纠正、影像增强、图像变换、结构信息提取、影像分类等技术，主要属于遥感图像处理的内容。

因此，以下主要介绍 GIS 中矢量数据的采集。GIS 中矢量数据的采集主要包括地图跟踪数字化和地图扫描数字化。

(1)地图跟踪数字化。跟踪数字化是目前应用广泛的一种地图数字化方式，是通过记录数字化板上点的平面坐标来获取矢量数据的。其基本过程是：将需数字化的图件(地图、航片等)固定在数字化板上，然后设定数字化范围、输入有关参数、设置特征码清单、选择数字化方式(点方式和流方式等)，就可以按地图要素的类别

分别实施图形数字化了。

由于跟踪数字化本身几乎不需要 GIS 的其他计算功能，所以跟踪数字化软件往往可以与整个 GIS 系统脱离开，因而可单独使用。

地图跟踪数字化时数据的可靠性主要取决于操作员的技术熟练程度，操作员的情绪会严重影响数据的质量。操作员的经验和技能主要表现在能选择最佳点位来数字化地图上的点、线、面要素，判断十字丝与目标重合的程度等能力。为了保持一致的精度，每天的数字化工作时间最好不要超过 6 小时。

GIS 中的地图跟踪数字化软件为了获取矢量数据应具有下列基本功能：

——图幅信息录入和管理功能。即对所需数字化的地图的比例尺、图幅号、成图时间、坐标系统、投影等信息进行录入和管理。这是所采集的矢量数据的数据质量的基本依据。

——特征码清单设置。特征码清单是指安放在数字化仪台面或屏幕上的由图例符号构成的格网状清单，每种类型的符号占居清单中的一格。在数字化时只要点中特征码清单区的符号所在的网格，就可知道所数字化要素的编码，以方便属性码的输入。地图跟踪数字化软件应能使用户方便地按自己的意愿设置和定义特征码清单。

——数字化键值设置。即设置数字化标识器上各按键的功能，以符合用户的习惯。

——数字化参数定义。主要是指系统应能选定不同类型的数字化仪，并确定数字化仪与主机的通讯接口。

——数字化方式的选择。主要是指选择点方式还是流方式等进行数字化。

——控制点输入功能。应能提示用户输入控制点坐标，以便于进行随后的几何纠正。

(2)地图扫描数字化。扫描数字化是目前较为先进的地图数字

化方式，但要实现完全自动化还要做大量艰巨的努力，目前所能提供的扫描数字化软件是半自动化的，还需做相当的人机交互工作。地图扫描数字化的基本思想是：首先通过扫描将地图转换为栅格数据，然后采用栅格数据矢量化的技术追踪出线和面，采用模式识别技术识别出点和注记，并根据地图内容和地图符号的关系，自动给矢量数据赋属性值。

根据目前的技术水平，大多数软件要对所扫描的彩色地图进行分版处理，通常分为黑版要素、水系版要素、植被要素和地貌要素，也可以直接对分版图进行扫描，然后由软件进行二值化、去噪音等处理，有时需要进行一些编辑，以保证自动跟踪和识别的进行。目前已有一些软件不需要对彩色地图进行分版处理，可直接在彩色地图数据上进行矢量化。在软件自动进行跟踪和识别时，仍需要进行部分的人机交互，如处理断线、确定属性值等，有时甚至要人工在屏幕上进行数字化。

与地图跟踪数字化相比，地图扫描数字化具有速度快、精度高、自动化程度高等优点，正在成为 GIS 中最主要的地图数字化方式。

地图扫描数字化的自动化程度高，但必须具有一些对扫描后的地图数据的预处理能力，同时，由于其最后结果同地图跟踪数字化的结果是相同的，因而还必须具有地图跟踪数字化所具有的一些功能。因此，其基本功能可描述为：

——地图扫描输入功能。即能使用各种扫描仪把地图扫描数字化为栅格数据。

——图像格式转换和图像编辑功能。能接受不同格式的栅格数据，并具有基本的图像编辑功能。

——彩色地图图像数据的分版功能。能够将所扫描的彩色地图图像分成不同要素版的图像数据，以便于跟踪和识别。对于可在彩色地图图像数据上直接矢量化的软件，不需要这部分功能。

——线状要素的矢量化功能。能够对线状要素进行细化、断线

修复、跟踪，也即具有自动提取线状要素中心线的功能。由于目前的自动化程度还不够高，有时需要进行人机交互，如在多条线的交叉点，需人机交互指明继续追踪的方向。

——点状符号和注记的自动识别功能。应该能对点状符号和注记字进行自动识别，但完全自动化目前仍有困难，因此，有时需要人工在屏幕上进行数字化。

——属性编码的自动赋值。应能对已数字化的要素自动根据其符号特征赋以相应的编码（包括等高线的高程）。这方面目前还需要较多的人机交互。

——图幅信息录入与管理功能。同地图跟踪数字化一样，地图扫描数字化也需要录入图幅信息，以便于管理和质量控制。

——要素编码设置功能。为了能进行属性编码的自动赋值，以及人机交互地进行属性编码赋值，都必须针对不同的要求进行地图要素的编码设置。

——控制点输入功能。为了进行数字化后的数据纠正，必须具有控制点输入功能。

2. 属性数据的采集

属性数据在GIS中是空间数据的组成部分。属性数据的录入主要采用键盘输入的方法，有时也可以辅助于字符识别软件。

当属性数据的数据量较小时，可以在输入几何数据的同时，用键盘输入；但当数据量较大时，一般与几何数据分别输入，并检查无误后转入到数据库中。

为了把空间实体的几何数据与属性数据联系起来，必须在几何数据与属性数据之间有一公共标识符，标识符可以在输入几何数据或属性数据时手工输入，也可以由系统自动生成（如用顺序号代表标识符）。只有当几何数据与属性数据有一共同的数据项时，才能将几何数据与属性数据自动地连接起来；当几何数据或属性数据没有公共标识码时，只有通过人机交互的方法，如选取一个空间实体，再指定其对应的属性数据表来确定两者之间的关系，同时

自动生成公共标识码。

当空间实体的几何数据与属性数据连接起来之后，就可进行各种 GIS 的操作与运算了。当然，不论是在几何数据与属性数据连接之前或之后，GIS 都应提供灵活而方便的手段以对属性数据进行增加、删除、修改等操作。

3. 空间数据的检核

(1)空间数据输入的误差。空间数据输入的误差通常可归结为以下几类：

——几何数据的不完整或重复。

——几何数据的位置不正确。

——比例尺不正确。

——变形。

——几何数据与属性数据的连接有误。

——属性数据错误。

(2)空间数据的检查。无论是地图跟踪数字化还是地图扫描数字化，都不可能完全正确，因此，必须进行空间数据的检查。常用的空间数据检查方法为：

——通过图形实体与其属性的联合显示，发现数字化中的遗漏、重复、不匹配等错误；

——在屏幕上用地图要素对应的符号显示数字化的结果，对照原图检查错误；

——把数字化的结果绘图输出在透明材料上，然后与原图叠加以发现错漏；

——对等高线，通过确定最低和最高等高线的高程及等高距，编制软件来检查高程的赋值是否正确；

——对于面状要素，可在建立拓扑关系时，根据多边形是否闭合来检查，或根据多边形与多边形内点的匹配来检查等；

——对于属性数据，通常是在屏幕上逐表、逐行检查，也可打印出来检查；

——对于属性数据还可编写检核程序，如有无字符代替了数字，数字是否超出了范围，等等；

——对于图纸变形引起的误差，应使用几何纠正来进行处理。

(3)图形显示和数据处理。为了对 GIS 所输入的数据进行显示和检查，GIS 应提供下列功能：

——符号设计与符号库建立功能。为了能以不同符号表示不同类型的矢量数据，必须具有符号设计与符号库建立功能。如新符号的创建、旧符号的修改等等。

——符号设置功能。即为每一类空间数据指定选用的符号，包括符号的形状、色彩、尺寸、图案等。

——注记配置功能。注记是地图上不可缺少的重要信息，也是数据检查的重要内容和参照信息。注记应确定其字体、大小、间隔、色彩、排列、旋转等，最重要的是确定其定位点。

——图形显示功能。应能将所采集的矢量数据，以符号化的方式显示在屏幕上，并能进行放大、缩小、漫游、分层显示等操作。

——查询功能。通过查询来发现问题，可以由几何数据查询其属性信息，也可由属性信息查询其空间数据。

——绘图输出功能。即通过绘图机把所数字化的地图再以符号化的形式输出，这是数据检查的基本方法。

由数据输入软件获取的图形数据在进入地理数据库之前还需进行一些图形处理，主要包括：

——几何纠正。这是为了纠正由纸张变形所引起的数字化数据的误差，直接关系到 GIS 数据的质量。几何纠正要以控制点的理论坐标和数字化坐标为依据来进行，最后应显示平差结果。

——投影变换。为了 GIS 地理数据库中空间数据的一致性，须将原图投影下的矢量数据转换为地理坐标或指定投影下的数据。

——图形接边。在相邻地图的接合处可能会产生裂隙。包括几何裂隙和属性裂隙。在自动接边无法处理时，需要人机交互进

行。

——图形编辑功能。矢量数据错漏的纠正很大程度上依赖于强大的图形编辑功能。图形编辑功能应能对点、线、面进行增加、删除、移动、修改(如线的连接、截断、属性编码的修改)等,并应具有良好的人机界面和较快的响应速度。

——自动拓扑功能。拓扑关系是强大的查询与分析功能的基础。自动拓扑是指在已矢量化的数据的基础上,自动建立起点、线、面的拓扑关系。

§5.3 空间数据标准

从技术的角度看,空间数据标准是指空间数据的名称、代码、分类编码、数据类型、精度、单位、格式等的标准形式。每个地理信息系统都必须符合有关的空间数据标准。空间数据标准涉及到复杂的科学理论和技术方法问题。如果只针对某一地理信息系统设计空间数据标准,并不困难;如果所建立的空间数据标准能为大家所承认,为大多数系统所接受和使用,就比较复杂和困难。空间数据标准的制定对于地理信息系统的发展具有重要意义,但目前空间数据标准的研究仍然落后于地理信息系统的发展。

目前我国已有一些与GIS有关的国家标准,内容涉及数据编码、数据格式、地理格网、数据采集技术规范、数据记录格式等。

1. 空间数据分类标准

在地理信息系统中,空间数据必须按统一的标准进行分类。通常应遵循以下原则:

(1)遵循已有的国家标准,以利于全国范围内的数据共享。

(2)遵循国务院有关部委以及军队正在使用的数据标准。

(3)遵循各领域中普遍使用和认同的数据标准。

(4)当各种数据标准相互矛盾时,应遵循由上面下的原则进行处理。

(5)制定新的数据标准时，应尽可能参考同类标准。

目前我国已有的与 GIS 有关的关于空间数据分类的国家标准如：

GB2260—95　《中华人民共和国行政区划代码》

GB13923—92　《国土基础信息数据分类与代码》

GB11708—89　《公路桥梁命名和编码规则》

GB14804—93　《1:500、1:1000、1:2000 地形要素分类与代码》等等。

空间数据的分类体系是设计数据标准的前提，而分类体系应考虑专业领域专家的意见，并根据地理信息系统的要求来制定，尽可能反映分类的合理性。

2. 空间数据交换标准

随着地理信息系统的发展，数据共享已越来越重要。由于空间数据模型的不同，空间数据的定义、表达和存储方式也不同，因而数据交换就不那么简单。空间数据交换的主要方式有：

(1)外部数据交换标准。这类标准通常是 ASCII 码文件，用户可以通过阅读说明书来直接读写这种外部数据格式。GIS 的外部数据交换格式通常包括矢量数据交换格式、栅格数据交换格式和数字高程模型交换格式。

(2)空间数据互操作协议。制定一套各方都能接受的标准空间数据操纵函数，通过调用这些函数以互相操作对方的数据。

(3)空间数据共享平台。采用客户机/服务器体系结构，各种 GIS 通过一个公共的平台在服务器存取所有数据，以避免数据的不一致性。

(4)统一数据库接口。在对空间数据模型有共同理解的基础上，各系统开发专门的双向转换程序，将本系统的内部数据结构转换成统一数据库的接口。

随着软硬件的发展，人们逐渐感受到了外部数据交换格式的不足，如自动化程度不高，速度较慢等，但它毕竟解决了不同 GIS

之间的数据转换问题。

虽然空间数据互操作协议比外部数据交换标准方便，但由于各种软件存储和处理空间数据的方式不同，空间数据的互操作函数又不可能很庞大，因此往往不能解决所有问题。

空间数据共享平台虽然是一个较好的思路，但现有的 GIS 软件各有自己的底层，因此目前难以实现。

统一数据库接口，首先要求对现实世界进行统一的面向对象的数据理解，这也是不易实现的。

因此，目前外部数据交换标准仍然是实现数据共享的主流方式。

§5.4 两种空间数据交换格式简介

1. 军用数字地图矢量数据交换标准简介

对于一幅图或一个区域，有 4 种文件：公用文件、属性文件、图元文件、地名文件。

(1)公用文件(*.OGY)。该文件以"OGY"为后缀，每幅地图有一个公用文件，记录数据的密级、要素层、标准等，以及图号、图名、比例尺、坐标范围、投影、坐标系等元数据，具体由文件头记录和元数据记录两部分组成。详细内容可参考该标准的文件。

(2)属性文件(*.? SX)。军用数字地图矢量数据交换标准的属性数据文件以"? SX"为后缀，用于记录点、线、面、复合要素的属性数据(如主码、识别码、描述码、参数码等)。

军用数字地图矢量数据根据有关标准分为 14 层(测量控制点、独立地物、居民地、交通等)，每一层有一个属性文件。文件后缀中的"?"为层的标识符，如 S 表示测量控制点，I 表示独立地物等。

属性文件的基本结果为：

- 文件头记录(记录数据的密级、要素层、标准等)
- 点状要素记录
  - 点状要素标识字符(P)
  - 点状要素个数

- 点状要素性数据
  - 编号(唯一编码)
  - 主码
  - 识别码
  - 描述码个数(描述码,…)
  - 参数码个数(参数码,…)
  - 类型(有向、无向、节点)
  - 地名数据记录号
- 线状要素记录
  - 线状要素标识字符(L)
  - 线状要素个数
  - 线状要素属性数据
    - 编号(唯一编码)
    - 主码
    - 识别码
    - 描述码个数(描述码,…)
    - 参数码个数(参数码,…)
    - 类型(曲线、折线)
    - 地名数据记录号
- 面状要素记录
  - 面状要素标识字符(A)
  - 面状要素个数
  - 面状要素属性数据
    - 编号(唯一编码)
    - 主码
    - 识别码
    - 描述码个数(描述码,…)
    - 参数码个数(参数码,…)
    - 类型(曲线、折线)
    - 地名数据记录号
- 复合要素记录
  - 复合要素标识字符(L)
  - 复合要素个数
  - 复合要素属性数据
    - 编号(唯一编码)
    - 主码
    - 识别码
    - 描述码个数(描述码,…)
    - 参数码个数(参数码,…)

- 类型(曲线、折线)
- 地名数据记录号
- 复合要素数(复合要素,…)
- 点状要素数(点状要素,…)
- 线状要素数(线状要素,…)
- 面状要素数(面状要素,…)

(3)图元文件(*.?ZB)。图元文件以"?ZB"为后缀,用于表示坐标数据和拓扑关系。每个要素层有一个图元文件。文件后缀中的"?"为层的标识符,如S表示测量控制点,I表示独立地物等。

- 文件头记录(记录数据的密级、要素层、标准等)
- 点状图元
  - 点状要素标识字符(P)
  - 点状要素个数
  - 点状要素性数据
    - 点状图元编号
    - 定位点(x1,y1,z1)
    - 方向点(x2,y2,z2)
    - 关联弧段图元数(弧段图元号,…)
- 弧段图元
  - 弧段图元标识字符(L)
  - 弧段图元个数
  - 弧段图元数据
    - 弧段图元编号
    - 定位点(x1,y1,z1),…,(xn,yn,zn)
    - 首、末结点(点状图元号)
    - 左、右面号(面状图元号)
- 面状(多边形)图元
  - 面状图元标识字符(A)
  - 面状图元个数
  - 面状图元数据
    - 面状图元编号
    - 内点(x1,y1)
    - 关联弧段图元数(弧段图元号,…)

(4)地名文件(*.NDM)。该文件以"NDM"为后缀,每幅地图有一个地名文件,记录地名数据。

- 文件头记录(记录数据的密级、要素层、标准等)

├─地名标识字符(N)
├─地名记录数
├─地名记录1
├─ ⋮
└─地名记录 n ┌─地名记录号
　　　　　　├─地名内容,…
　　　　　　├─ ⋮
　　　　　　└─定位点坐标

2. 我国矢量空间数据交换格式简介

对于一幅图或一个区域,只有一个文件(*.VPF),由6部分组成。

┌─文件头,包含基本数据,如图幅范围坐标维数、比例尺等。
├─要素类型参数
├─属性数据结构
├─几何数据
├─注记
└─属性数据

(1)文件头。文件头不限定字节数,以 HeadBegin 开始,以 HeadEnd 结束。文件分两类数据:一类是基本的必须的数据,一类是扩充的附加信息。主要存储本交换格式的标志、版本号、坐标单位、图幅范围、比例尺、投影、参考椭球、数据源说明等。

(2)要素类型参数。要素类型参数部分以 FeatureCodeBegin 开始,以 FeatureCodeEnd 结束。每个要素的内容为:

┌─要素类型编码
├─要素名称
├─几何类型(点、线、面、注记)
├─缺省颜色
├─属性表名
└─其他用户项

(3)属性数据结构。属性数据结构部分以 TableStructure-Begin 开始,以 TableStructureEnd 结束。每个属性表表示为:

——属性表名
——属性项个数
——属性项
—— ⋮
——属性项 ——属性项名
　　　　　——字段描述

(4)几何数据。把点、线、面、注记相对集中地放在一起。

点状要素以 PointBegin 开始,以 PointEnd 结束。每个点目标表示为:

——目标标识码
——要素类型编码
——层名
——点的特征类型(独立、有向、节点)
——(x1,y1,z1),(x2,y2,z2)
——关联线对象的个数(线对象标识号,…)(如果有拓朴关系时)

线状要素以 LineBegin 开始,以 LineEnd 结束。每个线目标表示为:

——目标标识码
——要素类型编码
——层名
——坐标说明(直接坐标或间接坐标)
——线的特征类型(折线、圆弧、椭圆、光滑曲线、B 样条)
——点数,坐标((x1,y1,z1),(x2,y2,z2),…)(如果是直接坐标时)
——子线(弧段)数,(线标识号 1,线标识号 2,……)(如果是间接坐标时)
——始节点,终节点,左面,右面(如果有拓扑关系时)

面状要素以 PolygonBegin 开始,以 PolygonEnd 结束。每个面目标表示为:

——目标标识码
——要素类型编码
——层名

- 标识点坐标(x,y)
- 坐标说明(直接坐标或间接坐标)
- (直接坐标时)
  - 多边型 1 的点数((x1,y1,z1),(x2,y2,z2),…)
  - 多边型 2 的点数((x1,y1,z1),(x2,y2,z2),…)
  - ⋮
  - 0
- (间接坐标时)线目标数(线目标标识号,…)(0 为岛的标识,负号为反向连接)

(5)注记。注记以 AnnotationBegin 开始,以 AnnotationEnd 结束。每个目标注记表示为:

- 目标标识码
- 要素类型编码
- 层名
- 字体、颜色、尺寸、间隔、…
- 注记内容
- 坐标点数,注记坐标与方向(X,Y,Φ),…)

(6)属性数据。属性数据以 AttributeBegin 开始,以 AttributeEnd 结束。由若干属性表组成,每个属性表的格式为:

- 属性表名
- 属性表内容
  - 目标标识码
  - 分隔符,属性项(字符串、整形、浮点、日期…)
  - ⋮
  - 分隔符+属性项
- 属性表结束标志(TableEnd)

§5.5 GIS 空间元数据

随着地理信息系统在社会各方面的发展,越来越多的地理学科和信息技术学科之外的个人、组织和机构也涉入到这一领域,开始生产、处理和修改数字地理信息。但是这些机构从各自的角度出发来发展空间数据,使得人们不知道存在什么样的数据、已有数据

的质量如何、以及怎样访问和使用这些数据成果。因此，迫切需要采取一定的办法来避免数据的重复性建设，同时协调不同数据部门之间的资源共享，这样随着地理空间数据集的数量、复杂性和多样性的增加，一个适应数据集共享的标准化规范——空间元数据，也就应运而生。

对空间元数据标准内容的研究，国际上主要有美国联邦数据委员会(FGDC)、国际标准化组织地理信息/地球信息技术委员会(ISO/TC211)和开放的GIS联盟(OGC)进行，它们从不同侧面对地理空间元数据进行了描述。国内近几年来已经提出了几个元数据标准，如《中国可持续发展信息共享 Metadate 标准》、《中国生态系统研究网络元数据标准》和《科学数据库元数据标准》等。

1. 空间元数据的定义及其作用

元数据(Metadate)：数据的数据，是关于数据和信息资源的描述性信息。图书馆的图书卡片就是关于所有书籍的简单的元数据，它记录了每本书的编号、题目、作者、关键字和出版日期等属性。

空间元数据(Geospatial Metadata)：地理的数据和信息资源的描述性信息。它通过对地理空间数据的内容、质量、条件和其他特征进行描述与说明，以便人们有效地定位、评价、比较、获取和使用与地理相关的数据。空间元数据是一个由若干复杂或简单的元数据项组成的集合。如果说地理空间数据是对地理空间实体的一个抽象映射，那么可以认为，空间元数据是对地理空间数据的一个抽象映射。空间元数据和地理空间数据是对地理空间实体不同层次的描述，是对地理信息的不同深度的表达。

综合起来，空间元数据主要有以下几个方面的作用：

(1)用来组织和管理空间信息、并挖掘空间信息资源，这正是数字地球的特点和优点所在。通过它可以在 Intranet 或 Internet 上准确地识别、定位和访问空间信息。

(2)帮助数据使用者查询所需空间信息。比如，它可以按照不同的地理区间、指定的语言以及具体的时间段来查找空间信息资源。

(3)组织和维护一个机构对数据的投资。通过空间元数据内容,可以充分描述数据集的详细情况,便于数据使用者得到数据的可靠性保证。同时,当使用数据引起矛盾时,数据提供单位也可以利用空间元数据维护其利益。

(4)用来建立空间信息的数据目录和数据交换中心。通常由一个组织产生的数据可能对其他组织也有用,而通过数据目录、数据代理机、数据交换中心等提供的空间元数据内容,用户便可以很容易地使用它们,达到空间信息的共享。

(5)提供数据转换方面的信息。通过空间元数据,人们便可以接受并理解数据集,并可以与自己的空间信息集成在一起,进行不同方面的分析决策,使地理空间信息实现真正意义上的共享,发挥其最大的潜力。

2. 空间元数据的分类

空间元数据可以按照所描述的对象分为三层:高层元数据对应数据库,中层元数据对应表,底层元数据对应数据项。各种元数据与描述地理实体的空间数据之间的关系如图 5—1 所示。

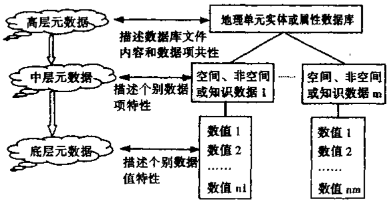


图 5—1 空间元数据层次结构

(1)高层元数据(数据集系列 Metadata)。高层元数据是指一

系列拥有共同主题、日期、分辨率以及方法等特征的空间数据系列或集合，它是用户用于概括性查询数据集的主要内容。在软件实现上，如果拥有数据集系列 Metadata 模块，则既可以使数据集生产者方便地描述宏观数据集，而且也可以使用户很容易地查询到数据集的相关内容，实现空间信息资源的共享。当然，要获取数据集的详细信息，还需要通过中层元数据来实现。

(2)中层元数据(数据集 Metadata)。数据集 Metadata 模块是整个 Metadata 标准软件的核心，它既可以作为数据集系列 Metadata 的组成部分，也可以作为后面数据集属性以及要素等内容的父 Metadata 数据集系列。在 Metadata 软件标准设计的初级阶段，通过该模块便可以全面反映数据集的内容。然而随着数据集的变化，为了避免重复记录元数据内容以及保持元数据的实时性，需要通过继承关系更新变化了的信息，这时元数据的层次性便显得异常重要。

(3)底层元数据(要素、属性的类型和实例 Metadata)。要素类型是指由一系列几何对象组成的具有相似特征的集合，比如数据集中的道路层、植被层等便是具体的要素类型；要素实例是具体的要素实体，它用于描述数据集中的典型要素。属性类型是用于描述空间要素某一相似特征的参数，如桥梁的跨度是一个属性类型；属性实例则是要素实例的属性，如某一桥梁穿越某一道路的跨度。该 Metadata 模块是元数据体系中详细描述现实世界的重要部分，也是未来数字地球中走向多级分辨率查询的依据。因此，我们通过数据集系列、数据集、要素类型等层次步骤，便可以逐级对地理世界进行描述，用户也可以按照这一步骤，能够沿网络获取详细的数据集内容信息。

在具体实现时，可以用面向对象的技术实现以地理单元为主线的层次结构元数据管理树，树的叶结点连接空间数据物理存储的"层"，面向地理单元的空间操作通过元数据管理树作用于图层。

3. 空间元数据的内容

为了便于不同系统之间的空间数据和空间元数据相互交换，许多机构和组织对空间元数据所要描述的一般内容进行层次化和范式化，指定出可供参考与遵循的空间元数据标准的内容框架。

空间元数据标准由两层组成，其中第一层是目录层，它所提供的信息主要用于对数据集信息进行宏观描述，适合在数字地球的国家级空间信息交换中心或区域以及全球范围内管理和查询空间信息时使用。第二层是空间元数据标准的主体，它由八个基本内容部分和四个引用部分组成，其中基本内容部分包括标识信息、数据质量信息、数据集继承信息、空间数据表示信息、空间参考系信息、实体和属性信息、发行信息、以及空间元数据参考信息等八个方面的内容，另外四个引用部分包括引用信息、时间范围信息、联系信息以及地址信息。它们之间的关系如图 5—2。

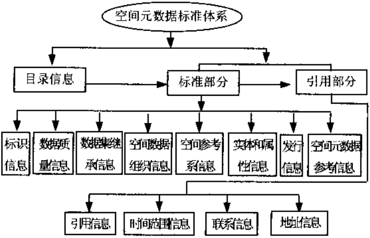


图 5—2　空间元数据内容的标准的组织框架

(1)标识信息。是关于地理空间数据集的基本信息。通过标识信息，数据生产者可以对有关数据集的基本信息进行详细的描述，诸如数据集的名称、作者信息、所采用的语言、数据集环境、专题分类、访问限制等，同时用户也可以根据这些内容对数据集有一个总

体了解。

(2)数据质量信息。是对空间数据集质量进行总体评价的信息。通过这部分内容用户可以获得有关数据集的几何精度和属性精度等方面的信息,也可以知道数据集在逻辑上是否一致以及它的完备性,这是用户对数据集进行判断以及决定数据集是否满足需要的主要判断依据。数据生产者也可以通过这部分内容对数据集质量评价的方法和过程进行详细的描述。

(3)数据集继承信息。是建立该数据集时所涉及的有关事件、参数、数据源等的信息,以及负责这些数据集的组织机构信息。通过这部分信息便可以对建立数据集的中间过程有一个详细的描述,比如当一幅数字专题地图的建立经过了航片判读、清绘、扫描、数字地图编辑以及验收等过程时,应对每一过程有一个简要描述,使用户对数据集的建立过程比较清晰,也使数据集每一过程的责任比较清楚。

(4)空间数据表示信息。是数据集中用来表示空间信息的方式的描述,如空间数据类型、空间数据结构、矢量对象描述、栅格对象描述等内容,它是决定数据转换以及数据能否在用户计算机平台上运行的必须信息。利用空间数据表示信息,用户便可以在获取该数据集后对它进行各种处理或分析了。

(5)空间参考系信息。是关于空间数据集地理参考系统与编码规则的描述,它是反映现实世界与地理数字世界之间关系的通道,诸如地理标识码参照系统、水平坐标系统、垂直坐标系统以及大地模型等。通过空间参考系中的各元素,可以知道地理实体转换成数字对象的过程以及各相关的计算参数,使数字信息成为可以度量和决策的依据。

(6)实体和属性信息。是关于数据集信息内容的信息,包括实体类型及其属性、属性值、域值等方面的信息。通过该部分内容,数据集生产者可以详细地描述数据集中各实体的名称、标识码以及含义等内容,也可以使用户知道各地理要素属性码的名称、含义

等。

(7)发行信息。是关于数据集发行及其获取方法的信息,包括发行部门、数据资源描述、发行部门责任、订购程序、用户订购过程以及使用数据集的技术要求等内容。通过发行信息,用户可以了解到数据集在何处,怎样获取、获取介质以及获取费用等信息。

(8)空间元数据参考信息。是关于空间元数据的标准、版本、现时性与安全性等方面的信息,它是当前数据集进行空间元数据描述的依据。通过该空间元数据描述,用户便可以了解到所使用的描述方法的实时性等信息,加深了对数据集内容的理解。

(9)引用信息。是引用或参考该数据集时所需的简要信息,它自己不单独使用,而是被基本内容部分的有关元素引用。它主要由标题、作者信息、参考时间、版本等信息组成。

(10)时间范围信息。是关于有关事件的日期和时间的信息,该部分是基本内容部分的有关元素引用时要用到的信息,它自己不单独使用。

(11)联系信息。是同与数据集有关的个人和组织联系时所需的信息,包括联系人的姓名、性别、所属单位等信息。该部分是基本内容部分的有关元素引用时要用到的信息,它自己不单独使用。

(12)地址信息。是同组织或个人通讯的地址信息,包括邮政地址、电子邮件地址、电话等信息。该部分是描述有关地址元素的引用信息,它自己不单独使用。

空间元数据是基于 Internet 的 GIS(即网络地理信息系统)必不可少的一部分,通过它可以了解 GIS 系统所提供地理空间数据的情况,如:有什么数据?数据质量如何?数据有哪些格式?以什么方式在哪儿可以得到数据?等等和数据有关的信息。通过这些信息可以实现地理空间数据的不同部门、不同专业领域的网络共享,避免因地理空间数据的重复收集、录入和处理导致的大量时间、人力和物力的浪费。

§ 5.6 空间数据的互操作和 OpenGIS 规范

空间数据的互操作是 UCGIS(The University Consortium for Geographic Information Science ,美国大学地理信息联盟)的优先研究领域之一。互操作不仅有理论性和前瞻性,而且在当前的 GIS 系统中有其实用性。目前 GIS 界已经制定了 OpenGIS 互操作规范,并且还在继续完善之中。

1. 空间数据的互操作

(1)概述。当前 GIS 虽然得到了广泛的应用,其使用范围涉及多学科和多部门,在资源管理、环境治理、预防灾害、区域规划、城市管理、科研、教育和国防等领域得到重要应用,但是 GIS 应用系统被认为是信息孤岛。应用部门在开发地理信息系统时通常依据本部门的空间数据格式,对地理数据的组织有很大的差异,使得在不同 GIS 软件上开发的系统间的数据交换存在困难,采用数据转换标准也只能部分解决问题。另外,不同的应用部门对地理现象有不同的理解,对地理信息有不同的定义,这使得领域间在共同协作中进行信息共享和交流存在障碍。这些问题归结为地理信息系统间的互操作问题。互操作是系统集成的基础。

各部门对 GIS 互操作有迫切的需要。首先是解决对基础数据的共享问题。各部门对建立的 GIS 应用系统,在有限的经济条件下,更着眼于本部门专业信息及模型的开发,而对基础数据及其他专业数据通过有偿共享而获得。其次,GIS 的应用部门对综合和集成有较强的要求,特别是社会发展、经济发展走向一体化,以及可持续发展的要求,需要多学科、多部门的协作。第三,GIS 正走向社会化,融于主流的信息技术,成为信息社会和信息基础设施的重要组成部分。如果 GIS 缺乏互操作性,这些就不可能实现。

互操作应包括哪些内容?以下是 GIS 互操作的几个定义。

“互操作地理信息处理”(Interoperable geoprocessing)是指数

字系统的这些能力：①自由地交换所有关于地球的信息，即所有关于地表上的、空中的、地球表面以下的对象和现象的信息；②通过网络协作运行能够操作这些信息的软件。概括为自由交换地理空间信息以及协作运行空间信息处理的软件。

所谓互操作，就是指异构环境下两个或两个以上的实体，尽管它们实现的语言、执行的环境和基于的模型不同，但它们可以相互通信和协作，以完成某一特定任务。这些实体包括程序、对象、系统运行环境等。

互操作在软件工业中意味着界面的开放。内部数据结构的公开发表允许 GIS 用户从不同的开发者那里集成软件构件，允许新的软件厂商用可竞争的产品进入市场，并可与已存在的构件互换。互操作也意味着系统间的自由数据交换，交换标准如空间数据转换标准(SDTS)对系统间的数据转换有重要的作用。

互操作方面的进步将由易用性来度量，由完成某一任务所需要的培训来表示。

(2) GIS 互操作现状。

——GIS 互操作实现的方法之一：OpenGIS 规范。为了发展地理信息的互操作技术，一方面要面向未来，着眼于新技术的应用和新型信息的处理，例如地理空间技术与多媒体技术的复合；另一方面也要照顾到传统 GIS 的技术积累。这就要求作为异构 GIS 系统之间互操作标准的规程不但应成为互操作的桥梁，而且还应构成 GIS 传统技术与未来技术之间沟通的桥梁。OGC 制定的 OpenGIS 规范就是为了满足 GIS 互操作需要而制定的。

通过这种规范的制定，OGC 为软件开发者提供了一个统一的设计和开发软件工具的框架，它使得用户能在分布式计算平台的网络环境下获取地理数据，由于采用了开放式技术来开发地理数据处理和服务工具，从而使得用户和不同的信息群在 Internet 和 Intranet 中能灵活地进行地理数据及处理的互操作。这种思路和方法给地理空间技术带来一个全新的变化，用数据、系统和机构之

间的具有互操作的地学处理代替了传统的孤立的地理信息技术，使得可以在数据库转换中保持数据的语义或尽可能地减少数据语义的丢失。

——可行的实现 GIS 互操作的方法。在选择互操作的实现方法时，所要掌握的一个原则是所有涉及地理数据和处理的机构必须能持久地维护和更新它们各自所拥有的 GIS 技术，并能使这些技术不断增值。这意味着在选择实现方法时必须对下列问题做出评估：

- 地理数据的可用性和兼容性；
- 所使用的地学处理系统的功能和可互操作性；
- 承担地学服务的机构的灵活性和创新性。

很显然，传统的孤立的地学处理系统不但费用昂贵，而且它们使用各自所独有的接口界面，以致于很难通用。由于这种不灵活性和不兼容性的接口使得互操作在实现时非常艰难。

因而，构件技术用于实现互操作是当前可行的方法。用新的互操作构件代替部分传统的 GIS。通过这种方法，新的构件能逐渐替代旧的孤立的系统，并改进新的客户端、新的服务体系结构和新的服务器功能。经过一步一步统一的过程，OpenGIS 规范将被实现，从而能设计出关于开放式地学处理系统的综合软件体系结构。

2. OpenGIS 规范

(1)概述。OpenGIS 是指开放的地学数据互操作规范(Open Geo－data Interoperability Specification，简称 OGIS)的最高层次，一般称为 OpenGIS 规范。

OpenGIS 规范是由开放地理信息系统协会(Open GIS Consortium，简称 OGC)制订的一系列开放标准和接口。OGC 由商业部门、政府机构、用户以及数据提供商等多个领域的成员组成。

OpenGIS 规范是 OGC 规范的最高层次，即有关地理处理互操作的完整定义。开放地理数据互操作规范(OGIS)是利用软件统

一表示地理数据和地理处理的规范系统。

OGC 的目的是通过信息基础设施，把地理空间数据和地理处理资源集成到主流的计算机技术中，促使可互操作的商业地理信息处理软件的广泛应用。OpenGIS 规范提供了地理信息及处理标准，按照这个规范开发的各个系统之间可以自由地交换地理信息和处理功能。

OGC 是唯一一个由各个成员组成的用于发展地理操作开放系统方法的组织。通过它的协调发展，OGC 在全球地理数据和地理处理标准方面已形成了很大的影响，并成功地促进了 OpenGIS 技术的发展。OpenGIS 技术是把地理处理与企业的分布计算体系和 Internet 组成在一起的一种技术。

(2)OpenGIS 规范的作用。我们即将面临一个分布式的网络计算时代。在这个时代，我们可以用一个统一的地理处理获取规范接口，即 OpenGIS 规范，使全球的地理信息连接在一起。通过把分布式计算、对象技术、中间件软件以及组件软件等引入地理处理世界，可以使任何计算环境或计算任务与空间数据联系在一起。同时我们看到，随着卫星传感器、GPS 技术的发展以及参与地理信息研究、生产、获取和使用的人员的增加，地理数据的增长速度在明显地加快。OpenGIS 规范的作用如图 5—3 所示，通过 OpenGIS 规范把商业部门、集成部门、用户、研究人员、数据提供商等连接到一起，通过必要的软件工具和通信技术，为各种用户提供对地理信息的共享和互操作。

(3)互操作地理处理的工作方式。非互操作起源于不同的软件厂商在定义地理对象的基本几何特征时使用了各自的内部结构，同时，当用户在使用每一商业 GIS 时，又作了不同目的的二次开发，这一切使得信息共享困难重重。地理信息团体中数百个分支机构中的每一个部门都有自己独特的模式模型，它们贯穿在信息表示技术的每一阶段，诸如数据字典、对象模式、几何规则、分类模式等等。

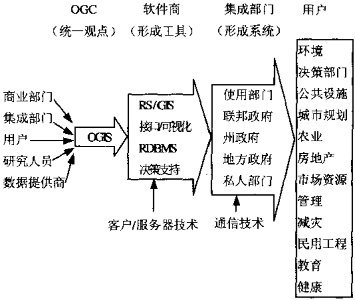


图 5—3 OpenGIS 规范的作用

当今的对象建模技术已经被用在多种分布式计算平台上，如何实现 OpenGIS 规范，OpenGIS 规范并没有提出一个具体的标准的实施模式。OpenGIS 规范的开放地理数据模型(OGM)包含一个众所周知的类型和结构集合。通过这一集合，OGC 技术委员会便可以表示任何地理模型，比如具体类型的集合构成一个地理模型(“类型”代表简单或复杂的数据结构，它们可以在特定数据集合中以一种具体“事例”或“结构”的方式重复多次，比如，一个字节便是最简单的一个类型，它可以对应一个具体的 ASCII 字符)。这些类型和结构被确定为接口的方式。一种实用的方法便是用这些众所周知的类型和结构表示信息，低层(软件引擎)和高层(信息群体)的模式模型可以用 OGM 的构件来表示。

对于信息团体有两种定义或观点。一种观点认为它是描述共享一个对象定义和语义的人类团体，特别在地理信息团体中，人们

为更好地实现他们的软件，便通过一些“行话”来满足他们的专业需求，产生了一个公共的模式模型。对“信息团体”一词的另一观点是指 OpenGIS 规范中描述的机制，该机制给出了地理信息在人类信息群体中的组织方式，并能通过网络进行异构地理数据和地理处理资源进行查询、集成、分类、交换等。

不同应用部门的信息团体对地理实体几何特征的描述方法也不一样，比如，在一个群体中“公路”是线，而在另一个中则是多边形。它们的属性也不尽相同，在一个群体中，小路的数量是关键要素；而在另一个中，路面类型则是考虑的主要因素。OGM 提供了一种描述几何特征和属性的唯一可行的统一的方法（在全球地理处理软件开发商之间达成了一致），这种方法是实现不同系统间通信接口的关键。

另外，依赖于 OGM 的 OpenGIS 规范中的 OpenGIS 服务体系，已集成了信息团体机制，确定了一种可以发行和查询这些定义及其所描述的数据类型的唯一可行和统一的方法。

第六章　空间数据的质量

在地理信息系统中，空间数据质量的好坏和空间数据标准化程度的高低，都直接影响到系统的经济效益和社会效益。近年来，中国在 GIS 的理论研究及产业化发展方面，有许多成功发展 GIS 的范例，同时也存在着一些问题。GIS 中的数据质量，即不确定性问题，是许多不成功系统的主要原因之一。本章通过介绍地理信息系统中不确定性分析的理论与方法，使更多的人了解到 GIS 中不确定的一面，从而在 GIS 的发展方面，可以更加全面、系统地考虑 GIS 的系统设计与开发。

§6.1　GIS 数据质量概述

1. GIS 数据质量的基本概念

(1)GIS 数据质量的定义

GIS 数据质量(GIS data quality)：GIS 数据产品满足一定使用要求的特性。主要包括数据源、点位精度、属性精度、要素完整性和属性完整性、数据逻辑一致性、数据现势性等。GIS 的数据质量通常用空间数据的误差和正确率来度量。

数据质量控制(data quality control)：为达到数据质量要求所采用的技术方法和手段的总称。就是对采集的空间数据确定误差的来源、性质和类型，提出度量误差的指标，分析误差在 GIS 空间操作中的传播机制，研究削弱误差对 GIS 产品质量影响的方法。

(2)研究 GIS 数据质量的重要意义

地理信息系统中的数据质量问题是一个关系到数据可靠性和

系统可信性的重要问题，与GIS系统的成败密切相关。由于现代GIS的先进技术，用户可以不管比例尺的大小、图形的精度而较容易地把来源不同的数据进行综合、覆盖和分析，其结果是误差的增加，导致不能决策，系统失败。

误差好比是GIS的孪生姐妹，建立GIS的过程就是和误差作斗争的过程。GIS产生的误差可以比喻成快速产生的高级垃圾，如对误差处理不当，GIS能以相当快的速度产生各种垃圾，而这种垃圾看起来似乎是精美无比的。如果不考虑GIS的数据质量，那么当用户发现GIS的结论与实际的地理状况相差大得惊人时，GIS产品就会在用户中立刻失去信誉。因此，GIS若要生存发展下去，必须花大力气从理论上研究GIS空间数据的质量问题。

从事GIS的人员，一定既要认识到在GIS中，误差是不能消除的，又要认识到误差是可以降低而且必须降低的。因此随时随地跟踪GIS误差理论的发展，并将其应用到实践中去，是我们每一个GIS工作者的任务。从长远的角度看，研究如何降低GIS系统中的误差，是GIS生存的生命线。

(3)GIS数据质量的国际标准

鉴于GIS数据质量的重要性，国际标准化组织地理信息/地球信息业技术委员会(ISO/TC211)从1995年开始立项研制第一批国际地理信息标准时，在其开展的20个项目中安排了两个与数据质量问题有关的项目，即：

ISO15046－13 地理信息——质量

ISO15046－14 地理信息——质量评价过程

——《地理信息——质量》标准。该标准旨在建立描述地理数据质量的原理和提供有关数据质量的模型。该标准定义数据质量信息的组成部分，包括数据质量定量元素、数据质量定量子元素、数据质量度量方法、数据质量非定量元素等。地理数据的生产者和用户都使用这一标准，生产者按照该标准说明地理数据的质量，用户则按照该标准确定地理数据是否满足其特定的应用需求。

地理信息数据质量的元素分为两大组成部分：

——数据质量的定量元素。它是数据生产者对照 GIS 数据生产规范的一组判断标准，度量其质量的好坏。数据质量定量元素包括若干子元素，每个子元素包括一种或几种度量方法。

——数据质量的非定量元素。它便于数据使用者根据所提供数据的目的、历史记录和使用信息，评价 GIS 数据对于某种特定应用的适用程度。

数据质量元素和子元素(即衡量空间数据质量的标准)：

· 完整性：如要素的完整性、属性的完整性等。

· 逻辑一致性：如属性一致性、格式一致性、拓扑一致性等。

· 位置精度：如绝对精度、像元位置精度、相对精度、形状的相似性等。

· 时间精度：如与时间有关的属性或属性类型误差等。

· 专题精度：如要素属性的连续值精度、有序值精度、额定值精度等。

——《地理信息——质量评价过程》标准。该标准主要是提供数字形式地理数据质量的评价过程框架。评价过程根据《地理信息——质量》标准定义的数据质量模型确定数据质量评价内容，建立评价过程框图，将数据质量评价结果报告作为数据质量描述数据(metadata 即元数据)的一部分。按照数据生产者的生产规范和用户对数据质量的要求，确定 GIS 数据满足应用需求的质量等级。

数据质量评价内容：

基于《地理信息——质量》标准中定义的数据质量组成部分，数据质量评价内容包括地理信息数据集、要素类型、要素、要素关系、属性类型、属性等。

数据质量评价过程：

该标准提出的地理信息数据质量评价过程如图 6—1 所示。数据生产者和用户在对 GIS 数据集进行质量评价时应遵循这一框架流程。

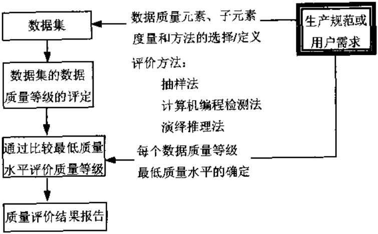


图 6－1　数据质量评价过程

数据质量评价方法：

直接评价法。用计算机程序自动检测。某些类型的错误可以用计算机程序自动发现，数据中不符合要求的数据项的百分率或平均质量等级也可由计算机程序算出。例如，可以检测文件格式是否符合规范、编码是否正确、数据是否超出范围等。当采用计算机程序自动检测时，用于检测的程序的名称、算法和其他参考信息应当在数据质量报告中说明，以使评价结果可信。基于随机抽样的检测方法。抽样检验可以对产品的质量管理提出可靠的信息，是质量控制的基本手段。传统的百分比抽样在检验数字产品的质量时有不足之处，在复杂情况下，需要在有足够地理空间数据抽样检测经验的统计学家参加下设计抽样方法。当应用基于抽样的评价方法时，评价过程必须作为质量报告的一部分予以简要说明。

间接评价法。间接评价法是基于外部知识或信息进行演绎推理来确定数据集的质量等级或符合程度。用于演绎推理的外部知识或信息有用途、数据历史记录、数据源的质量、数据生产方法和系统的信息以及误差传递模型等。用于演绎推理的方法和外部知识或信息应当在质量报告中阐述，以提高演绎推理的可信性。

非定量描述法(或综合法)。数据集、要素类型或属性类型的总体质量可以通过对数据质量的组成部分的质量评价结果进行综合来确定。在综合过程中,应考虑这些组成部分出错率的重要性。综合评价方法应当在数据质量评价报告中说明。

(4)空间数据的误差类型

Burrough(1986)曾系统地分析了 GIS 中的误差。他将误差分为:①明显误差;②源于自然或原始量测值的误差;③源于数据处理的误差。这三组误差中,第一组误差是十分明显而易于查出的;第二组误差相对比较复杂而在与数据接触时可以查出;第三组误差是在执行某一处理时才会出现,这种误差最难于量测,因为误差不仅与所使用的数据有关,而且与所使用的算法、数据结构有关。

——明显误差

明显误差包括:

· 数据年龄:大多数 GIS 使用已有的数据,而这种数据往往是过时的。使用的数据越旧,出现误差的可能性就越大。

· 地图比例尺:地图比例尺越大,表示的内容越详尽,一般来讲,精度也越高。

· 观测值的密度及其分布模式:有些专题图(如土壤图)是对一系列观测值的内插得来的。在建立数字地面模型时,内插误差较量测误差要大得多。有两个因素可影响其精度,即观测值的密度及其分布模式。

——源于自然变量或原始量测值的误差(又称为源误差)

主要是指数据采集和录入中产生的误差,包括:

· 遥感数据:摄影平台、传感器的结构及稳定性、信号数字化、光电转换、分辨率等。

· 测量数据:人差(对中误差、读数误差、平分误差等)、仪差(仪器不完善、缺乏校验、未作改正等)、环境影响(气候、气压、温度、磁场、信号干扰、风、照明度等)。

· 属性数据:属性数据的录入、数据库的操作等。

· GPS 数据:发射信号的精度、接收机精度、定位方法、信号处理算法等。

· 制图:控制点精度,编绘、清绘、制图综合等的精度。

· 数字化精度:纸张变形、数字化仪精度(定点误差、重复误差、分辨率)、操作员的技能、采样点密度、要素的图上宽度等。

—— 处理过程中产生的误差。GIS 对空间数据进行处理时产生误差,例如在下列处理中产生的误差就是处理误差。

· 几何纠正;

· 坐标变换;

· 几何数据的编辑;

· 属性数据的编辑;

· 空间分析(如多边形叠置等);

· 图形化简(如数据压缩);

· 数据格式转换;

· 矢量、栅格数据的相互转换;

· 计算机中的数字误差(如舍位误差)。

(5)GIS 中的误差传播

误差传播是指对有误差的数据,经过变换、处理后生成的 GIS 产品也存在着误差。误差传播在 GIS 中可归结为三种方式。

—— 代数关系下的误差传播:这是指对有误差的数据进行代数运算后,所得结果的误差。

—— 逻辑关系下的误差传播:即指在 GIS 中对数据进行逻辑交、并等运算所引起的误差传播,如叠置分析时的误差传播。

—— 推理关系下的误差传播:这是指不精确推理所造成的误差。

2. 研究 GIS 数据质量的常用方法

(1)建立误差分析体系。误差分析体系包括误差源的确定、误差的鉴别和度量方法、误差传播模型的建立以及控制和削弱误差对 GIS 产品影响的方法。传统的概率统计仍是建立误差分析体系

的理论基础。但是，必须根据GIS操作运算的特点对经典的概率统计理论进行扩展和补充。

(2)敏感度分析法。一般而言，精确确定GIS数据的实际误差非常困难。为了从理论上了解输出结果如何随输入数据的变化而变化，可以通过人为的在输入数据中加上扰动值来检验输出结果对这些扰动值的敏感程度。然后根据适合度分析，由置信域来衡量由输入数据的误差所引起的输出数据的变化。

为了确定置信域，需要进行地理敏感度测试，以便发现由输入数据的变化引起输出数据变化的程度，即敏感度。这种研究方法得到的并不是输出结果的真实误差，而是输出结果的变化范围。对于某些难以确定实际误差的情况，这种方法是行之有效的。

在GIS中，敏感度检验一般有以下几种：地理敏感度、属性敏感度、面积敏感度、多边形敏感度、增删图层敏感度等。敏感度分析法是一种间接测定GIS产品可靠性的方法。

(3)尺度不变空间分析法。地理数据的分析结果应与所采用的空间坐标系统无关，即为尺度不变空间分析，包括比例不变和平移不变。尺度不变是数理统计中常用的一个准则，一方面能保证用不同的方法得到一致的结果，另一方面可在同一尺度下合理地衡量估值的精度。

也就是说，尺度不变空间分析法使GIS的空间分析结果与空间位置的参考系无关，以防止由基准问题而引起分析结果的变化。

(4)Monte Carlo实验仿真。由于GIS的数据来源繁多，种类复杂，既有描述空间拓扑关系的几何数据，又有描述空间物体内涵的属性数据。对于属性数据的精度往往只能用打分或不确定度来表示。对于不同的用户，由于专业领域的限制和需要，数据可靠性的评价标准并不相同。因此，想用一个简单的、固定不变的统计模型来描述GIS的误差规律似乎是不可能的。在对所研究问题的背景不十分了解的情况下，Monte Carlo实验仿真是一种有效的方法。

Monte Carlo 实验仿真首先根据经验对数据误差的种类和分布模式进行假设，然后利用计算机进行模拟试验，将所得结果与实际结果进行比较，找出与实际结果最接近的模型。对于某些无法用数学公式描述的过程，用这种方法可以得到实用公式，也可检验理论研究的正确性。

(5)空间滤波。获取空间数据的方法可能是不同的，既可以采用连续方式采集，也可采用离散方式采集。这些数据采集的过程可以看成是随机采样，其中包含倾向性部分和随机性部分。前者代表所采集物体的实际信息，而后者是由观测噪声引起的。

空间滤波可分为高通滤波和低通滤波。高通滤波是从含有噪声的数据中分离出噪声信息；低通滤波是从含有噪声的数据中提取信号。例如经高通滤波后可得到一随机噪声场，然后用随机过程理论等方法求得数据的误差。

对 GIS 数据质量的研究，传统的概率论和数理统计是其最基本的理论基础，同时还需要信息论、模糊逻辑、人工智能、数学规划、随机过程、分形几何等理论与方法的支持。

3. 研究 GIS 数据质量的主要理论

在 GIS 领域中人们已经研究了许多处理 GIS 中误差的模型与方法。然而，到目前为止还没有提出一个处理 GIS 数据质量问题的一般性理论。但是，现有的若干理论可用于处理 GIS 数据质量的若干侧面问题，这些理论是：概率论、模糊数学、证据数学理论以及空间统计理论。这些理论在处理 GIS 数据质量问题时具有互补性，而且每一理论只可以处理某几方面的问题。

(1)概率论。概率论可用于处理由于随机误差而产生的数据质量问题。在概率论中，空间数据的误差被描述成给定某些观测值的条件下某一假设为真的条件概率。该假设的条件概率表示了一个概率从 0 到 1 区间的定量描述。例如在遥感分类中，误差是由条件概率表示出来的，它指出在划分某一个像素为某一类别时的概率。

在条件概率公式中，假设先验概率可以正确地确定，这在许多

应用中是难以达到的。这里没有一个概率值质量的描述指标，即本模型不能提供其可靠性的度量指标。

(2)模糊数学。模糊数学是由 Zadeh(1965)提出的由计算机处理不精确概率的一种理论。

在 GIS 中有时要处理如下问题 $\mu_A(O)\in[0,1]$：一个空间物体对一个集合的归属关系 $O\in A$(物体 O 属于集合 A)。集合 A 的定义可能是模糊的，在 GIS 中这可能意味着某一物体的自然边界的定义不清楚。这个模糊性可能在空间域、属性域、时间域、或(遥感影像的)频谱域，例如在空间域中的"城市区域"，属性域的"湿地"。对一个模糊集合，对于全集中的每一个元素(O)，需给出一个隶属函数值以表示该元素对于集合 A 的隶属度。

模糊数学的优点在于该理论在进行与自然语言有关空间查询方面具有较强的潜在应用可能。它的弱点是没有一个像概率论那样经过严格证明的过程。

(3)证据数学理论。证据数学理论是对传统概率论的一个扩展。在证据数学理论中，某一假设为真的概率量测值由支持某一假设的可获得的证据的概率来代替。证据理论是基于可信函数和可能函数所确定的一个区间，可信函数度量已有的证据对假设支持的程度，可能函数量测根据已有的证据不能否定假设的程度。可信函数表示了证据支持假设的最低程度，而可能函数则表达了证据支持假设的最高程度。证据数学理论可用于解决决策问题：O∈A (物体 O 属于集合 A)。

证据数学理论的优点在于其为概率论的更一般性表达。它的弱点是不能就矛盾证据或不同假设之间具有微弱的支持问题提出解决办法，对该理论值得信服的解释尚未给出。

(4)空间统计理论。统计学是用有序的模型描述无序事件的一门不确定性理论。Cressie 将空间数据分为三大类：地理统计数据、栅格数据以及点数据。他提出了一个通用模型，该模型适用于上述三种数据类型。

设 $S\in R^n$ 是一个 n 维欧几里德空间的一个一般数据。设一个潜在随机数据 z(s)在空间位置 s。现在假设 s 在一个索引集合 $D\in R^n$ 中,因此产生多维随机矢量场(或随机过程)

$$\{z(s);s\in D\}$$

根据 D 的定义不同可以有不同的空间数据(地理统计数据、栅格数据或点数据)。

传统的非空间模型是空间模型的一个特例,空间统计较之非空间统计是一个更一般性的理论。空间统计理论可应用于描述栅格属性数据,也可用于描述线段的不确定性。

§6.2 GIS 数据质量的评价

1. 数据采集中数据质量的评价

GIS 中数据采集的方法通常可分为直接方法和间接方法两种。直接方法是指直接从野外实地采集或利用摄影测量方法来获取 GIS 数据,间接方法是指从已有的地图或其他图件上进行采集。

直接方法获取的数据受人差、仪器误差、环境误差等的影响,这些误差均有相应的技术规范进行严格的限制,以保证采集到的数据达到精度和可靠性的指标。间接方法获取的数据中,除了含有直接方法中的误差外,还有展绘控制点的误差、编绘的误差、制图综合的误差、地图制印的误差、数字化的误差等。

地图数字化是获取矢量数据的主要方法之一,也是 GIS 中的重要误差源,是 GIS 数据质量研究的重点之一。在地图数字化中,原图固有误差和数字化过程中引入的误差是两个主要的误差源。下面对地图数字化的数据误差作一分析。

(1)地图固有误差的来源和类型。在地图的固有误差中,除了含有直接方法获取数据的误差外,至少还存在下列误差:

——控制点展绘误差。展绘控制点是成图的第一步。当对地

图的精度要求不高时，该项误差可不考虑。

——编绘误差。通常点状特征的编绘精度优于线状特征的编绘精度，即使都是线状特征，如果分辨率和要素宽度不同，编绘精度也不同。

——绘图误差。绘图误差是在绘图过程中产生的，其误差范围为 0.06～0.18mm。

——综合误差。综合误差的大小取决于特征的类型和复杂程度，又取决于采用的制图综合方法，如取舍、移位、夸大等，因此，综合误差极难量化。

——地图复制误差和分版套合误差。这些都是地图印刷中产生的误差，如地图复制误差的均方差为 0.1～0.2mm。

——绘图材料的变形误差。地图一般印在纸上，随着温度和湿度的变化，纸张的尺寸也会变化。由于纸张在印刷时温度升高，纸张长度会伸长 1.5%，宽度会伸长 2.5%；而当纸张干燥和冷却后，其长度和宽度又分别收缩 0.5%和 0.75%。因此，在地图印刷完成后，图纸在长、宽方向上的净伸长分别约为 1.0%和 1.75%。

——特征的定义。自然界中的许多特征并无明确的界限。例如，海岸线的位置、森林的边界等，但在地图上却有明确的位置。因此，特征定义会引起特征位置的某些不确定性。

——地图投影引起的误差。地图的比例尺是指真实的主比例尺。对 Lombert 正形投影而言，主比例尺只有沿着标准纬线时才是不变的，在标准纬线的内、外都发生改变。因此，在将地图数字化时，必须利用适当的比例尺因子进行修正。如果是从不同的地图上采集信息，必须了解地图的投影方式是否一致。

由于制图过程中各种误差间的关系以及图纸尺寸的稳定性不能确定，因此很难准确地评价原图的固有误差。假定总误差与上述各种误差间存在线性关系，则可由误差传播定律计算总误差。

(2)数字化的误差。目前的地图数字化方式主要有跟踪数字化和扫描数字化两种。数字化的精度主要受数字化仪的精度、数字化

方式、操作员的水平、数字化软件的算法等的影响，常采用下列方法进行评价。

·自动回归法。该方法是从统计学角度出发，用随机分析的方法来对数字化过程中的误差进行模型化。建立的模型应能对数字化图及其变量的误差影响进行评估，并且应能对各种数字化精度标准进行比较。

在对线划进行跟踪数字化的过程中，每隔一定时间和距离就记录一次坐标值，因此可以认为这些数据是序列相关的。即某一点误差的大小，除受该点本身的影响外，还受前一点误差的影响。在统计学上，这样的误差序列构成了一个随机序列，以数字化线上的点到该线真实位置上点的垂线距离为变量，应用自动回归方法，并且对每项回归得到的均值进行逐项移动变换，最后得出误差范围。

由于跟踪数字化不仅是一个随机序列，而且是一个时间序列，因此可用数理统计中的时间序列分析法来确定数字化的误差。

·ε—Band(误差带)法。该方法是一种类似于 GIS 中建立缓冲区操作的限定误差方法，即在一条数字化线的两侧，各定义宽为 ε 的范围，作为该数字化线的误差带，也就是用 ε 值来定义和说明误差的影响范围，以及处理多边形叠置等的误差。该方法适用于任何类型的 GIS 数据，关键是如何给出合理的 ε 值。国外多采用根据数据误差的统计性质，选择一个条件概率函数来定义该 ε 值。也可以采用二次多项式逼近的方法给出较合理的 ε 值。

·对比法。所谓对比法就是把数字化后的数据，用绘图机绘出，与原图叠合，选择明显地物点进行量测，以确定误差。除了位置精度外，属性精度、完整性、逻辑一致性等也可用对比法进行对照检查。发现错漏记入检查手簿，根据缺陷等级给出缺陷扣分值，最后按扣分值进行质量评定。

2. 数据处理中数据质量的评价

在 GIS 的数据处理中，几何纠正、坐标变换、格式转换等的计算，除了受计算机字长的影响外，在理论上可以认为是无误差的，

因此，数据处理过程中的主要误差集中在与应用直接相关的处理中，下面举几个例子说明。

(1)数字高程模型(DEM)的精度。数字高程模型(DEM)的数据来源是多种多样的，建立 DEM 的技术也不一样。常用的方法是利用解析立体测图仪从立体航空像对上测得高程，或者是利用数字化的地图等高线进行内插获得。此外，地面测量、声纳测量、雷达测量等数据也可作为 DEM 的数据源。

DEM 的精度主要受原始资料的精度(采样密度、测量误差、地形类别、控制点数目等)和内插的精度(内插方法、地形类型、原始数据的密度等)的影响。在研究 DEM 的精度时，一般都假定已排除了粗差的影响，因为 DEM 的粗差难以探测。

对 DEM 内插精度的估算方法有多种，但结论是相同的，即 DEM 的内插精度主要受原始采样点的采样密度的影响，与不同的插值方法的关系不很大。但在 DEM 精度评定的标准方面、地貌逼真度方面、DEM 的粗差探测等方面仍没有得到圆满的解决。

(2)矢量数据栅格化的误差。矢量数据栅格化的误差可分为属性误差和几何误差两种。

在矢量数据转换为栅格数据后，栅格数据中的每个像元只含有一个属性数据值，它是像元内多种属性的一种概括。例如，在陆地卫星图像上，每个像元对应的地面面积为 80m×80m，像元的属性值是像元内各地物发射量的平均值。如果像元内有一部分物体的反射率很高，既使占像元的面积比例很小，对像元属性值的影响也很大，从而导致分类错误，且损失一些其他有用信息。因此，像元越大，属性误差越大。

几何误差是指在矢量数据转换成栅格数据后所引起的位置的误差，以及由位置误差引起的长度、面积、拓扑匹配等的误差。几何误差的大小与像元的大小成正比。其中矢量数据表示的多边形网用像元逼近时会产生较严重的拓扑匹配问题。

关于拓扑匹配误差问题，Frolov 和 Maling(1969)提出的最早

的估算方法是考虑一个像元被一条实际边界线二等分时引起的误差问题。而后,Goodchild(1980)重新研究了这个问题。

如果假设边界线为一条随机穿过像元的直线,则对每个由边界线切割的像元 i,把其切除部分的平均面积定义为误差方差,其估算公式为:

$$\hat{\sigma}_i^2 = aS^4 \tag{6—1}$$

其中,S 是正方形像元的边长,a 为常数。Frolov 和 Maling 计算的 a 为 0.0452,后来 Goodchild 建议改为 0.0619。

多边形面积误差的估算是按全部多边形边界像元的误差和来计算的。如果一个多边形的边界像元有 m 个,则其误差方差为:

$$\hat{\sigma}^2 = maS^4 \tag{6—2}$$

边界像元的个数 m 可按多边形的周长来计算,且可简单地按下式计算:

$$m = \sqrt{N} \tag{6—3}$$

其中 N 为多边形内所包含的像元总数。

Switzer(1975)提出了估计矢量数据栅格化精度的另一种方法。其分析的前提是:假设误差仅仅是由栅格化引起的,不考虑观测误差。该方法的基本思想是:假设存在一幅理想的矢量地图,图上不同属性的制图单元由很细的线分开;对理想地图进行观测采样得到一幅具有规则格网的栅格地图,把这两幅图进行叠置比较。虽然理想地图是根本不存在的,但在这一假设下,可以提供一种仅利用栅格地图本身来估算矢量数据栅格化精度的方法。该方法的具体计算过程从略。

(3)空间数据扫描矢量化的误差。空间数据扫描矢量化的方法有两种:一是对特征明显、规律的要素(如等高线)采用自动定位方式;二是对图形复杂不易用算法实现栅格矢量转换的要素采用交互定位方式。由于第一种方法是采用算法自动获得的,它的误差是容易推算的。在这里只探讨采用交互方式获取矢量数据的误差。

地图经扫描仪扫描后可转化为一幅二值图像,在对地图要素

进行采集时，得到的首先是其在图像坐标系中的图像坐标(x，y)，要取得地图各要素(称为采样点)的矢量坐标(X_D，Y_D)须进行如下变换

$$\begin{cases} X_D = a_{11} + a_{12}x + a_{13}y + a_{14}xy \\ X_D = a_{21} + a_{22}x + a_{23}y + a_{24}xy \end{cases} \quad (6—4)$$

式中各变换参数对同一幅地图上的各点是固定不变的，可以根据四个图廓点(称为定向点)的图像坐标和相应的大地坐标来求得。

综合考虑定向点和采样点的量测误差的影响，提高点的量测精度和扫描分辨率是提高栅格数据矢量化精度的主要手段。提高扫描分辨率虽能提高精度，但图像数据的存储量也会随之大大增加，分辨率增加1倍相应的数据量将增加3倍，由于受到计算机内存大小和数字化时视窗大小的限制，分辨率的提高会给数字化作业过程增加麻烦。因此，在保证地图图像具有一定分辨率(如取250～500DPI)的前提下，采用自动对中量测方法量测定向点和采样点，就成为提高栅格数据矢量化精度的主要手段。

自动对中分为定向点自动对中和采样点自动对中。通过实验可以知道，定向点的精度对数字化精度有较大影响。就1:5万地形图而言，只要定向点采用自动对中方法求得，采样点即使在250DPI图像上直接量测，数字化结果也能满足控制点误差不超过图上0.2mm和其他采样点误差不超过图上0.4mm的规范要求。

(4)多边形叠置产生的误差。多边形叠置是GIS中常用的空间分析方法，但却会产生拓扑匹配误差、几何误差和属性误差。

多边形叠置误差计算的思路是：先计算单幅图或单层图的误差，再计算叠置图的误差。因此，下面仅简要讨论单数据层的叠置问题。

多边形叠置往往是不同类型的地图、不同的图层，甚至是不同比例尺的地图进行叠置，因此，同一条边界线往往是不同的数据，这样在叠置时必然会出现一系列无意义的多边形。所叠置的多边

形的边界越精确，越容易产生无意义的多边形。这就是拓扑匹配误差。

多边形叠置所形成的多边形的数量与原多边形边界的复杂程度有关。将两个分别含有 V_1 和 V_2 个顶点的多边形叠置后将产生至少 3 个，至多 V_1+V_2+1 个多边形。如果多边形之间具有统计独立性时，产生中等数量的多边形；如果是高度相关的，则产生大量无意义的多边形。

多层叠置产生的无意义多边形实际上相当于矢量多边形栅格化引起的面积匹配误差。面积匹配误差因数字化精度的提高而减少，虽然无意义多边形的个数增多。

不管是用人机交互的方法把无意义的多边形合并到大多边形中，还是根据无意义多边形的临界值，自动合并到大多边形中，以及用拟合后的新边界进行叠置，都会产生几何误差，既新边界可能会偏离已制图的边界位置（或真实位置）。为了保证人们习惯上认为重要的边界线的精度，如境界、河流、主要道路等，处理时应对这些边界上的点加权使他们能尽可能地不被移动。

除了几何误差，实际上每个进行叠置的多边形本身的属性就是有误差的，因为属性值是分类的结果（如把植被分为不同的类别），而分类就会产生误差。多幅图的叠置会使误差急剧增加，以至使叠置出的结果不可信。

§6.3 基于等高线的 DEM 内插算法评价方法

GIS 中有许多数学模型和算法，在算法的运算过程中也会不同程度地产生误差。因此，GIS 中的数学模型和算法也需要进行评价。这里以基于等高线的 DEM 内插算法评价方法为例，说明一种评价 GIS 中算法的方法。

1. DEM 内插算法评价的思路

DEM 数据的质量不仅依赖于原始数据的精度，而且与内插

方法有关。在DEM数据质量的研究方面，已经有了许多研究成果。以往的研究重点主要集中在数据获取方法、采样密度、结果精度和地形类型之间的关系；或者对某种内插方法对DEM数据质量的可能影响进行理论研究；还有人直接以DEM数据为基础，根据地形的自相关性进行DEM误差检测及修正方法的研究；还有对以DEM为数据源的空间分析和计算的精度进行研究的。但是，关于如何评价各种DEM内插算法的实际效果的研究却很少。

许多DEM内插方法的目的是使高程误差的均方差达到最小，这一准则被普遍的运用于DEM的质量评估。相反，有些算法是为了保持地形的结构特征，有时甚至不惜降低高程的精度。本文的研究目的是不考虑各种内插算法的理论依据，面只考虑算法的实际应用效果(以及算法的软件编程的优劣)，并以可视化的方式将评价结果显示出来。

DEM的误差源主要有两种：一是对实际地形表面采样所引起的误差，即原始数据的误差；另一种是数据重采样引起的误差，即由DEM内插算法引起的误差。为了评价DEM内插算法引起的误差，需要屏蔽原始数据的误差。因此，DEM内插算法评价的基本思路是：

——建立一个严格的数学曲面模型，由该模型生成等值线，作为原始数据。由于都是严格的数学运算，不会引入任何误差(计算误差忽略不计)。这样就可忽略原始数据的误差。

——由等值线数据运用各种不同的内插算法生成DEM数据。这里产生了DEM内插算法引起的误差。

——将生成的DEM数据与相应的原始数学曲面模型进行比较，就可以找到DEM内插算法引起的误差。

——将误差进行统计分析，并用可视化的方式显示出来。

该项研究对10多种DEM内插算法进行了评价，由于篇幅所限，以下面4种算法(取之于有关软件)为例：

- 膨胀法
- 四方向曲线光滑加权法

·单点移面法　　　　·二元样条法

2. **数学曲面的建立和等值线的生成**

(1)数学曲面模型的建立。为了对DEM内插算法进行较全面的评价,本文建立了两类曲面模型。一类是简单的数学曲面,另一类是可调节的数学曲面。

——简单的数学曲面模型。在研究中,使用了五种简单的数学曲面:半球面、反置半球面、旋转抛物面、正四棱锥表面、圆锥表面。这几种曲面上点的坡度、坡向分布都具有一定的代表性,可以用来对DEM内插算法进行评估。下面以半球面和旋转抛物面为例,给出曲面方程和梯度的公式,取1000 * 1000为最大的试验区域,以(500,500)为曲面的水平中心点。

半球面方程:

$$(x-500)^2+(y-500)^2+(R-z)^2=R^2 \quad (0<R<500)$$

$$|grad_{半球面}|=\frac{\sqrt{R^2-z^2}}{z} \qquad (6-5)$$

注:x,y,z为三维坐标,x,y水平方向,z竖直方向,R为半球面半径(预先给定)。

在z的定义域内,梯度随z的减小而增大,梯度的变化速度随z的减小而增大,梯度变化域为0°~∞°,坡度从0°~90°。

旋转抛物面方程:

$$\frac{(x-500)^2+(y-500)^2}{2P}=(H-z)$$

$$(H>0, P>0, 2PH<250000)$$

$$|grad_{半球面}|=\frac{\sqrt{2P(H-z)}}{P} \qquad (6-6)$$

注:x,y,z为三维坐标,x,y水平方向,z竖直方向,H,P为旋转抛物面参数(预先给定)。

在z的定义域内,梯度随z的减小而增大,梯度的变化速度随z的减小而增大,但没有半球面的变化速度快,梯度变化域为0~

$(2Ph)^{1/2}/P$。

— 自定义数学曲面。自定义数学曲面由已知的 M * N 个格网点控制，并且可以调节。曲面由双二次 Hermite 插值得到。由于篇幅有限，双二次插值函数表达式具体见《样条函数与计算几何》(孙家昶，1982，科学出版社)。曲面的每个子矩形是由四块一阶光滑连接的曲面拼接而成的，而整个矩形域是由 4MN 块双二次多项式所组成，整个曲面函数连续，且两个方向的一阶导数连续。

自定义数学曲面可以模拟具有代表性的地形，如：平原地形、丘陵地形、山地地形，在原有的 DEM 数据基础上构造曲面，然后进行 DEM 内插算法的分析，可检验 DEM 内插算法对于各种地形的适应性。

(2)等值线的生成。等值线是水平面与数学曲面相交而成的曲线，取样过程中用离散点表示连续曲线，其取样密度由曲线的复杂程度确定。

简单数学曲面(半球面、旋转抛物面等)的等值线通过曲面与水平截面相交即可得到。自定义曲面与水平截面相交所得曲线均为不规则曲线，生成等值线的取样过程比较复杂，但基本方法是一样的。

3. DEM 内插算法的评价指标

对于 DEM 内插算法的评估很难提出一个通用的评价标准，本文主要通过下列评价指标来评价算法的优劣。

(1)DEM 数据与原始曲面对比。由等值线插值生成 DEM 会出现许多不正常的现象，如"梯田现象"、"斑纹现象"等。所以，通过 DEM 格网数据点的高程分布，运用可视化的手段，可看出内插算法的优劣。

(2)中误差和最大误差。此评价指标是对格网点的高程值不符合真值的情况进行检验。

中误差：
$$\sqrt{\sum_{i=1}^{imax}\sum_{j=1}^{jmax} d_{i,j}^2 / imax \cdot jmax} \qquad (6-7)$$

式中：$d_{i,j}$为第 i 列、第 j 行的格网点高程值与其对应曲面上高程值的差值，imax 为格网的列的总数、jmax 为行的总数。中误差是内插生成的 DEM 数据格网点相对于真值的偏离程度。这一准则被普遍的运用于 DEM 的质量评估。

最大误差反映了算法得到的格网点的高程值不符合真值的最坏情况。

(3)精确全局误差。精确全局误差可描述内插引入的曲面逼近误差。精确全局误差可通过计算 DEM 与原始数学曲面之间的"豪斯道夫距离"(Hausdorf distance)实现。

点 x 和一个 R^n(n 维欧几里德空间)中的点集 Y 之间的欧几里德距离 d(x,Y)定义为点 x 到集合 Y 中所有点的距离中的最小距离：

$$d(x,Y)=\inf_{y\in Y} d(x,y) \qquad (6—8)$$

其中 d(x,y)表示 R_n中的两点之间的欧几里德距离。由此可定义集合 X 到集合 Y 的单侧豪斯道夫距离(one－sided Hausdorff distance)dE(X,Y)为对于 x∈X，d(x,Y)的最大值：

$$d_E(X,Y)=\sup_{x\in X} d(x,Y) \qquad (6—9)$$

如果原始模型 Σ 与简化模型 Σ′ 之间的单侧豪斯道夫距离 d_E(Σ,Σ′)小于值 ε，则有：

显然，利用单侧豪斯道夫距离 dE(Σ,Σ′)即可描述简化模型 Σ 相对于原始模型 Σ 的误差。

$$\forall\ x\in\Sigma,\text{存在 } y\in\Sigma',\text{满足 } d(x,y)<\varepsilon$$

4. 评价结果分析

(1)DEM 数据与原始曲面对比。用不同内插算法生成的格网 DEM 数据生成三维透视图，即可看出格网 DEM 数据点的高程分布，对 DEM 内插算法就有了一个基本的评价。对半径为 500 的半球面用前面所述的 4 种内插算法生成格网 DEM 数据(格网密度为 200＊200、等高距为 10)，所生成的三维透视图如图 6—2 所示。

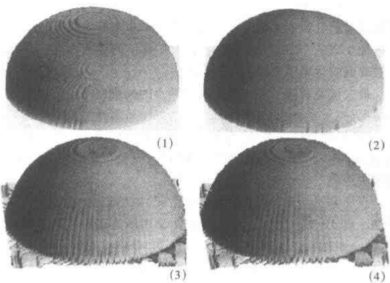


图 6—2 (1)膨胀法　　(2)四方向曲线光滑加权法,
(3)单点移面法　　(4)二元样条法

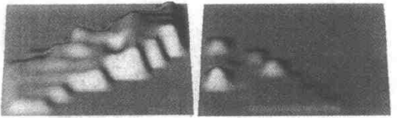
(2)中误差和最大误差。表 6—1 是四种算法对半球面生成的 DEM 格网数据(格网密度为 200 * 200、等高距为 10)进行检验的结果。表中数据的单位与等高距的单位相同。所形成的曲面如图 6—3 所示。

最大误差往往发生在坡度变化最大的区域,或者是对特殊情况没有处理。例如,对于半球面,最大误差往往发生在平面与半球面的交界处(如膨胀法、单点移面法),某些算法的最大误差是由于其对平面没有处理而形成的。

(3)精确全局误差计算。对于旋转抛物面,取等高距为 10,格网密度为 200×200,高程区间为[0,500],计算出的 4 种内插算法生成的 DEM 数据对应的单侧豪斯道夫距离的结果见表 6—2。表中数据的单位与等高距的单位相同。

表 6—1　DEM 格网数据的中误差和最大误差

数学曲面 / 插值方法	自定义曲面一 高程区间 [0,468]		自定义曲面二 高程区间 [0,243]		旋转抛物面 高程区间 [0,500]		半球面 高程区间 [0,500]	
	最大误差	中误差	最大误差	中误差	最大误差	中误差	最大误差	中误差
膨胀法	−38.0	8.92	10.4	8.04	10.2	3.68	48.4	5.4
四方向曲线光滑加权法	−10.0	6.32	10.0	1.32	3.36	0.75	19.7	0.98
单点移面法	−64.0	8.62	−21.0	6.97	−34.7	8.88	85.8	11.19
二元样条法	−63.0	7.89	−21.0	3.15	−34.7	8.89	84.9	11.18



自定义曲面一　　　　自定义曲面二

图 6—3　自定义曲面略图

(4)DEM 内插算法的误差分布。在标准数学曲面的立体显示的基础上，使用格网点上的误差数据代替曲面 DEM 高程进行分层设色，可以看出误差在数学表面上的分布情况。图 6—4 为半径为 500 的旋转抛物面(等高距 10、格网密度为 200×200)用膨胀算法形成的 DEM 误差分布情况(越暗误差越大)。

(5)DEM 内插算法对于地形坡度变化的适应性分析。DEM 内插算法需要适应不同的地貌类型(如平原地形、丘陵地形、山地地形等)。在对不同内插算法进行误差计算后，根据不同的坡度进行统计，得到坡度与内插算法的关系，从而从一个侧面对各种内插

算法的地形适应性进行评价。

表 6－2

内插算法	单侧豪斯道夫距离
膨胀法	10.26
四方向曲线光滑加权法	11.92
单点移面法	34.70
二元样条法	34.70

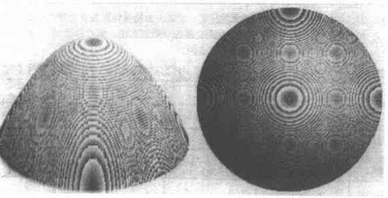


图 6 4 膨胀算法在旋转抛物面 DEM 上的误差分布

选择半径为 500 的半球面(等高距 10、格网密度为 200×200)，将坡度分为 10 级，9 度为一级，统计出的 4 种内插算法在不同坡度区域的中误差如表 6－3 所示。表中数据的单位与等高距的单位相同。

可以发现，几乎所有内插算法的中误差都随着坡度的增加而增大。单点移面法、二元样条法对于小坡度和大坡度情况处理不好(见表 6－3)。

5. 结论

通过以上分析，可以直观地评价各种 DEM 内插算法的优劣

和特点。用上述方法对算法进行评价，可以综合考虑算法的理论依据和算法的具体实现，对于在 DEM 建立的实际作业时如何选择一种较好的算法(软件)，具有实际的意义。通过对文章中的 4 种 DEM 生成算法(软件)的分析，四方向曲线光滑加权法较好。

另外，这种屏蔽原始数据的误差，单独研究算法形成的误差的方法，对于 GIS 数据质量研究中数据处理误差的研究具有重要的参考价值。

表 6－3　坡度与内插算法的关系(中误差)

坡度＼算法	膨胀法	四方向曲线光滑加权法	单点移面法	二元样条法
0－9°	1.670	0.585	3.652	3.638
9－18°	2.859	0.578	1.137	1.094
18－27°	3.061	0.625	1.163	1.175
27－36°	3.107	0.662	1.180	1.178
36－45°	3.242	0.690	1.270	1.226
45－54°	3.551	0.716	1.431	1.425
54－63°	3.997	0.761	3.267	3.260
63－72°	4.626	0.827	2.100	2.002
72－81°	5.843	0.948	3.672	3.614
81－90°	12.89	1.737	20.79	21.06

第七章　空间数据的处理

§7.1　空间数据的图形编辑

图形编辑是纠正数据采集错误的重要手段，其基本的功能要求是：具有友好的人机界面，即操作灵活、易于理解、响应迅速等；具有对几何数据和属性编码的修改功能，如点、线、面的增加、删除、修改等；具有分层显示和窗口功能，便于用户的使用。图形编辑的关键是点、线、面的捕捉，即如何根据光标的位置找到需要编辑的要素，以及图形编辑的数据组织。下面分别作简要介绍。

1. 点的捕捉

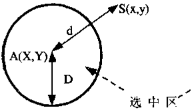


图 7—1　点的捕捉原理

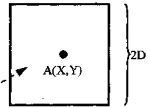


图 7—2　点捕捉的改进

图形编辑是在计算机屏幕上进行的，因此首先应把图幅的坐标转换为当前屏幕状态的坐标系和比例尺。设光标点为 S(x,y)，图幅上(图 7—1)某一点状要素的坐标为 A(X,Y)，则可设一捕捉半径 D(通常为 3～5 个像素，这主要由屏幕的分辨率和屏幕的尺寸决定)。若 S 和 A 的距离 d 小于 D 则认为捕捉成功，即认为找到的点是 A，否则失败，继续搜索其他点。d 可由下式计算：

$$d=\sqrt{(X-x)^2+(Y-y)^2} \qquad (7—1)$$

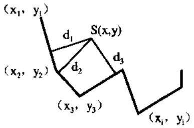
但是由于在计算 d 时需进行乘方运算，所以影响了搜索的速度，因此，把距离 d 的计算改为：

$$d=\max(|X-x|,|Y-y|) \qquad (7-2)$$

即把捕捉范围由圆改为矩形，这可大大加快搜索速度，如图 7—2 所示。

2. 线的捕捉

设光标点坐标为 S(x,y)，D 为捕捉半径，线的坐标为(x_1,y_1)，(x_2,y_2)，…(x_n,y_n)。通过计算 S 到该线的每个直线段的距离 d_i(如图 7—3 所示)，若 $\min(d_1,d_2,\cdots d_n-1)<D$，则认为光标 S 捕捉到了该条线，否则为未捕捉到。在实际的捕捉中，可每计算一个距离 d_i 就进行一次比较，若 $d_i<D$，则捕捉成功，不需再进行下面直线段到点 S 的距离计算了。



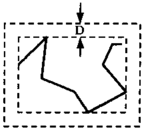


图 7--3 线的捕捉原理　　　图 7—4 线捕捉的改进

为了加快线捕捉的速度，可以把不可能被光标捕捉到的线以简单算法去除。如图 7—4，对一条线可求出其最大最小坐标值 Xmin，Ymin，Xmax，Ymax，对由此构成的矩形再向外扩 D 的距离，若光标点 S 落在该矩形内，才可能捕捉到该条线，因而通过简单的比较运算就可去除大量的不可能捕捉到的情况。

对于线段与光标点也应该采用类似的方法处理。即在对一个线段进行捕捉时，应先检查光标点是否可能捕捉到该线段。即对由

线段两端点组成的矩形再往外扩 D 的距离，构成新的矩形，若 S 落在该矩形内，才计算点到该直线段的距离，否则应放弃该直线段，而取下一直线段继续搜索。

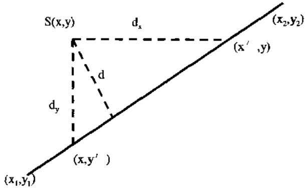


图 7—5 点到线段距离的计算方法

如图 7—5 所示，点 S(x,y)到直线段(x_1,y_1)，(x_2,y_2)的距离 d 的计算公式为：

$$d=\frac{|(x-x_1)((y_2-y_1)-(y-y_1)(x_2-x_1)|}{\sqrt{((x_1-x_2)^2+(y_1-y_2)^2}} \qquad (7—3)$$

可以看出计算量较大，速度较慢，因此可按如下方法计算。即从 S(x,y)向线段$(x_1,y_1)(x_2,y_2)$作水平和垂直方向的射线，取 dx，dy 的最小值作为 S 点到该线段的近似距离。由此可大大减小运算量，提高搜索速度。计算方法为：

$$x'=\frac{(x_2-x_1)(y-y_1)}{y_2-y_1}+x_1$$

$$y'=\frac{(y_2-y_1)(x-x_1)}{x_2-x_1}+y_1$$

$$dx=|x'-x|$$

$$dy=|y'-y|$$

$$d=\min(dx,dy)$$

3. 面的捕捉

面的捕捉实际上就是判断光标点 S(x,y)是否在多边形内，若在多边形内则说明捕捉到。判断点是否在多边形内的算法主要有垂线法或转角法，这里介绍垂线法。

垂线法的基本思想是从光标点引垂线(实际上可以是任意方向的射线)，计算与多边形的交点个数。若交点个数为奇数则说明该点在多边形内；若交点个数为偶数，则该点在多边形外(图 7－6)。

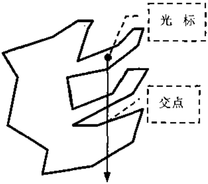


图 7－6 垂线法的思想

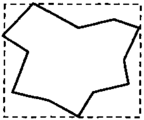


图 7—7 通过外切矩形提高捕捉速度

为了加速搜索速度，可先找出该多边形的外切矩形，即由该多边形的最大最小坐标值构成的矩形，如图 7—7。若光标点落在该矩形中，才有可能捕捉到该面，否则放弃对该多边形的进一步计算和判断，即不需进行作垂线并求交点个数的复杂运算。通过这一步骤，可去除大量不可能捕捉的情况，大大减少了运算量，提高了系统的响应速度。

在计算垂线与多边形的交点个数时，并不需要每次都对每一线段进行交点坐标的具体计算。对不可能有交点的线段应通过简单的坐标比较迅速去除。对图 7—8 所示的情况，多边形的边分别为 1～8，而其中只有第 3、7 条边可能与 S 所引的垂直方向的射线相交。即若直线段为 $(x_1,y_1)(x_2,y_2)$ 时，若 $x_1 \leqslant x \leqslant x_2$，或 $x_2 \leqslant x \leqslant x_1$

时才有可能与垂线相交，这样就可不对 1,2,4,5,6,8 边进行继续的交点判断了。

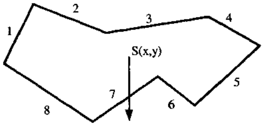


图 7—8 计算垂线与多边形的交点个数

对于 3、7 边的情况，若 $y>y_1$ 且 $y>y_2$ 时，必然与 S 点所作的垂线相交(如边 7)；若 $y<y_1$ 且 $y<y_2$ 时，必然不与 S 点所作的垂线相交。这样就可不必进行交点坐标的计算就能判断出是否有交点了。

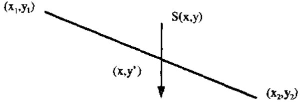


图 7—9 判断垂线是否与直线段有交点

对于 $y_1 \leqslant y \leqslant y_2$ 或 $y_2 \leqslant y \leqslant y_1$，且 $x_1 \leqslant x \leqslant x_2$ 或 $x_2 \leqslant x \leqslant x_1$ 时，如图 7—9。这时可求出铅垂线与直线段的交点 (x,y')，若 $y'<y$ 则是交点；若 $y'>y$，则不是交点；若 $y'=y$ 则交点在线上，即光标在多边形的边上。

以上都是一些提高面捕捉算法的常用技术。

4. 图形编辑的数据组织

GIS 中的空间数据通常是分层存取的，通常可分为控制点、独立地物、居民地、境界、水系、地貌、植被等。在进行图形编辑时，需确定在什么数据层(或哪几个数据层)进行操作，以便对选定数据

层的数据进行编辑。

无论空间数据库是用数据库管理还是用文件管理,都必须要为图形编辑的实现提供空间数据的读取、存储等基本功能。

由于GIS空间数据所涉及的地理区域较大,空间数据量是海量,因而每次编辑都针对全部空间数据进行是不现实的,因为这样的查询和编辑操作所需的时间是不能忍受的,所以,都需要采用建立索引的方法,索引通常是分层建立的,主要方法有四叉树索引和格网索引。

例如格网索引法就是把空间数据所涉及的地理区域划分为m×n的矩形格网。格网的大小可根据空间数据的数据量等情况确定。然后,用格网对所有空间数据进行扫描,以每个格网为索引单元,记录通过该格网的点、线、面。这样在查询时,根据光标的坐标可迅速算出处在哪个格网中,通过索引就可知道该格网中的点、线、面的内容,只对这些内容进行搜索和编辑,就可大大加快图形编辑的响应速度。

建立了索引文件后的图形编辑,不仅要修改原始的空间数据,而且要修改相关的索引文件。例如在增加了一条链后,不仅要在空间数据库中增加这条链的标识码、坐标串等数据,而且要在这条链经过的网格索引中增加这条链的信息。

在对空间数据进行删除操作时,通常不直接删除空间数据库中的相关数据,而只是在相应的索引文件中作一标志,只有在重新整理数据库时,才进行真正的删除。这种方法一方面加快了编辑的响应速度,另一方面也可在必要时恢复执行删除操作的数据。

在对建立了拓扑关系的矢量数据进行图形编辑时,往往会破坏原有的拓扑关系,这时需要拓扑重构。也可以先对图形编辑所涉及的局部区域进行拓扑重构,然后与原区域进行相关处理,以获取全图的拓扑关系数据。

§7.2 空间数据的几何纠正

在图形编辑中，只能消除数字化产生的明显误差，而图纸变形产生的误差难以改正，因此要进行几何纠正。几何纠正常用的有高次变换、二次变换和仿射变换。

1. 高次变换

$$\begin{cases} x' = a_1x + a_2y + a_{11}x^2 + a_{12}xy + a_{22}y^2 + A \\ y' = b_1x + b_2y + b_{11}x^2 + b_{12}xy + b_{22}y^2 + B \end{cases} \tag{7—4}$$

其中 A、B 代表二次以上高次项之和。上式是高次曲线方程，符合上式的变换称为高次变换。在进行高次变换时，需要有 6 对以上控制点的坐标和理论值，才能求出待定系数。

2. 二次变换

当不考虑高次变换方程中的 A 和 B 时，则变成二次曲线方程，称为二次变换。二次变换适用于原图有非线性变形的情况，至少需要 5 对控制点的坐标及其理论值，才能求出待定系数。

3. 仿射变换

$$\begin{cases} x' = a_1x + a_2y + a_3 \\ y' = b_1x + b_2y + b_3 \end{cases} \tag{7—5}$$

仿射变换是使用最多的一种几何纠正方式，只考虑到 x 和 y 方向上的变形，仿射变换的特性是：

- 直线变换后仍为直线；
- 平行线变换后仍为平行线；
- 不同方向上的长度比发生变化。

对于仿射变换，只需知道不在同一直线上的三对控制点的坐标及其理论值，就可求得待定系数。但在实际使用时，往往利用 4 个以上的点进行纠正，利用最小二乘法处理，以提高变换的精度。

误差方程为：

$$\begin{cases} Q_x = X - (a_1x + a_2y + a_3) \\ Q_y = Y - (b_1x + b_2y + b_3) \end{cases} \quad (7-6)$$

其中,X、Y 为已知的理论坐标。

由 Q_x^2 最小和 Q_y^2 最小的条件可得到两组法方程：

$$\begin{cases} a_1\sum x + a_2\sum y + a_3n = \sum X \\ a_1\sum x^2 + a_2\sum xy + a_3\sum x = \sum xX \\ a_1\sum xy + a_2\sum y^2 + a_3\sum y = \sum yX \end{cases}$$
$$\begin{cases} b_1\sum x + b_2\sum y + b_3n = \sum Y \\ b_1\sum x^2 + b_2\sum xy + b_3\sum x = \sum xY \\ b_1\sum xy + b_2\sum y^2 + b_3\sum y = \sum yY \end{cases} \quad (7-7)$$

其中 n 为控制点个数,x,y 为控制点坐标,X,Y 为控制点的理论值,a_1,a_2,a_3,b_1,b_2,b_3 为待定系数。

通过上述法方程就可求得仿射变换的待定系数。

§7.3 空间数据的自动拓扑

矢量数据拓扑关系在空间数据的查询与分析中非常重要,矢量数据的自动拓扑算法是 GIS 中的关键算法之一,下面介绍其实现的基本步骤和要点。矢量数据自动拓扑的步骤可分为以下几步：

1. 链的组织

找出在链的中间相交(图 7—10(1)),而不是在端点相交(图 7—10(2))的情况,自动切成新链;把链按一定顺序存储,如按最大或最小的 x 或 y 坐标的顺序,这样查找和检索都比较方便,然后把链按顺序编号。

2. 结点匹配

结点匹配是指把一定限差内的链的端点作为一个结点,其坐标值取多个端点的平均值,如图 7—11。然后,对结点顺序编号。



图 7-10　链的重新组织

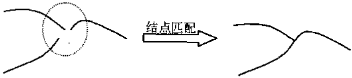


图 7—11　结点匹配原理

3. 检查多边形是否闭合

检查多边形是否闭合可以通过判断一条链的端点是否有与之匹配的端点来进行。如图 7—12，弧 a 的端点 P 没有与之匹配的端点，因此无法用该条链与其他链组成闭合多边形。多边形不闭合的原因可能是由于结点匹配限差的问题，造成应匹配的端点未匹配，或由于数字化误差较大，或数字化错误，这些可以通过图形编辑或重新确定匹配限差来确定。另外，还可能这条链本身就是悬挂链，不需参加多边形拓扑，这种情况下可以作一标记，使之不参加下一阶段拓扑建立多边形的工作。

4. 建立多边形

建立多边形是矢量数据自动拓扑中最关键的部分，由于其算法比较复杂，所以，先介绍几个基本概念，然后介绍其实现的过程。

(1)概念

——顺时针方向构多边形。

所谓顺时针方向构多边形是指多边形是在链的右侧。如图 7—13(1)，多边形在面内；图 7—13(2)，多边形在面外。

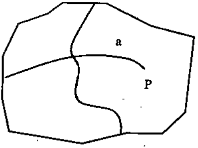
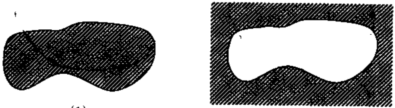


图 7—12　检查多边形是否闭合



(1)　(2)

图 7—13　顺时针方向构多边形

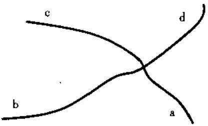


图 7—14　最靠右边的链

——最靠右边的链。

最靠右边的链是指从链的一个端点出发，在这条链的方向上右边的第一条链。如图 7—14，a 的最右边的链为 b。找最靠右边的

链可通过计算链的方向和夹角实现。

——多边形面积的计算。

设构成多边形的坐标串为(x_i,y_i)，$i=1,2,\cdots,N$，则多边形的面积 A 可用如下公式求出：

$$A=\frac{1}{2}\sum_{i=1}^{n}x_i(y_{i+1}-y_{i-1}) \qquad (7-8)$$

其中，当$i=n$时，$y_{n+1}=y_1$；当$i=1$时，$y_0=y_n$。根据该公式，当多边形由顺时针方向构成时，面积为正；反之，面积为负(图 7-15)。



(1) (2)

图 7—15 正、负多边形

(2)建立多边形的基本过程

①顺序取一个结点为起始结点，取完为止；取过该结点的任一条链作为起始链。

②取这条链的另一结点，找这个结点上，靠这条链最右边的链，作为下一条链。

③是否回到起点：是，已形成一多边形，记录之，并转④；否，转②。

④取起始点上开始的，刚才所形成多边形的最后一条边作为新的起始链，转②；若这条链已用过两次，即已成为两个多边形的边，则转①。

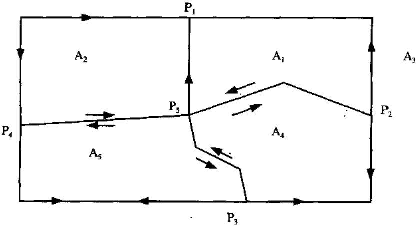


图 7—16 建立多边形的过程

例如，对图 7—16，建立多边形的过程为：

——从 P_1 结点开始，起始链定为 P_1P_2；从 P_2 点算起，P_1P_2 最右边的链为 P_2P_5；从 P_5 算起，P_2P_5 最右边的链为 P_5P_1。所以，形成的多边形为 $P_1P_2P_5P_1$。

——从 P_1 结点开始，以 P_1P_5 为起始链，形成的多边形为 $P_1P_5P_4P_1$。

——从 P_1 开始，以 P_1P_4 为起始链形成的多边形为 $P_1P_4P_3P_2P_1$。该多边形为包络多边形，面积为负。

—— 这时 P_1 为结点的所有链均被使用了两次，因而转向下一个结点 P2，继续进行多边形追踪，直至所有的结点取完。共可追踪出五个多边形，即 A_1、A_2、A_3、A_4、A_5。

5. 岛的判断

岛的判断即指找出多边形互相包含的情况，也即寻找多边形的连通边界。

根据上述追踪多边形的方法，单多边形即由单条链或由多条链顺序构成的，不与其他多边形相交的多边形，如图 7—17 被追踪了两次，因为每条链必须使用两次，所以，多边形的一个面积是为

正的，另一个是为负的。如果一个多边形包含另一多边形，则必然是面积正的多边形包含面积负的多边形。所以，解决多边形包含问题的步骤为：

①计算所有多边形的面积。

②分别对面积为正的多边形和面积为负的多边形排序。

③从面积为正的多边形中，顺序取每个多边形，取完为止。若负面积多边形个数为0，则结束。

④找出该多边形所包含的所有面积为负的多边形，并把这些面积为负的多边形加入到包含它们的多边形中，转③。

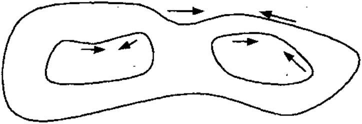


图7—17 正多边形包含负多边形

注意，由于一个面积为负的多边形只能被一个多边形包含，所以，当面积为负的多边形被包含后，应去掉该多边形，或作一标志。所以，当没有面积为负的多边形时，也应停止判断。

在该算法中，找出正面积多边形包含的负面积多边形是关键，其基本过程可描述为：

——找出所有比该正面积多边形面积小的负面积多边形。

——用外接矩形法去掉不可能包含的多边形。即负面积多边形的外接矩形不和该正面积多边形的外接矩形相交或被包含时，则不可能为该正面积多边形包含。

——取负面积多边形上的一点，看是否在正面积多边形内，若在内，则被包含；若在外，则不被包含。

6. 确定多边形的属性

在追踪出每个多边形的坐标后，经常需确定该多边形的属性。

如果在原始矢量数据中，每个多边形有内点，则可以把内点与多边形匹配后，把内点的属性赋于多边形。由于内点的个数必然与多边形的个数一致，所以，还可用来检查拓扑的正确性。如果没有内点，则必须通过人机交互，对每个多边形赋属性。

§7.4 矢量数据的压缩

矢量数据压缩的目的是删除冗余数据，减少数据的存贮量，节省存贮空间，加快后继处理的速度。下面介绍几种常用的矢量数据的压缩算法，以及它们之间的异同点。

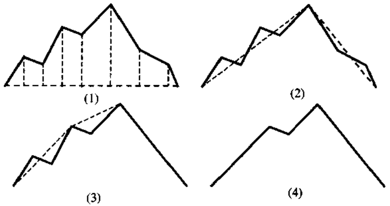


图 7—18 道格拉斯—普克法示意图

1. 道格拉斯——普克法（Douglas—Peucker）

基本思路是（图 7—18）：对每一条曲线的首末点虚连一条直线，求所有点与直线的距离，并找出最大距离值 dmax，用 dmax 与限差 D 相比：

若 dmax<D，这条曲线上的中间点全部舍去；

若 dmax≥D，保留 dmax 对应的坐标点，并以该点为界，把曲线分为两部分，对这两部分重复使用该方法。

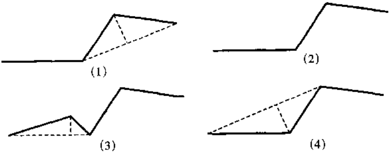


图 7—19　垂距法示意图

2. 垂距法

垂距法的基本思路是(图 7—19):每次顺序取曲线上的三个点,计算中间点与其它两点连线的垂线距离 d,并与限差 D 比较。若 d<D,则中间点去掉;若 d≥D,则中间点保留。然后顺序取下三个点继续处理,直到这条线结束。

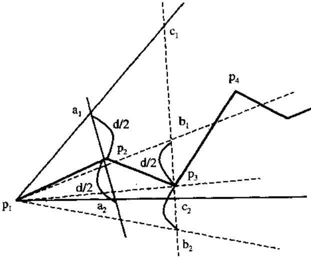


图 7—20　光栏法原理图示

3. 光栏法

光栏法的基本思想是(图 7—20)：定义一个扇形区域，通过判断曲线上的点在扇形外还是在扇形内，确定保留还是舍去。设曲线上的点列为$\{p_i\}$，i＝1，2，…，n，光栏口经为 d，可根据压缩量的大小自己定义，则光栏法的实施步骤可描述为：

(1)连接 p_1 和 p_2 点，过 p2 点作一条垂直于 p_1p_2 的直线，在该垂线上取两点 a_1 和 a_2，使 $a_1p_2=a_2p_2=d/2$，此时 a_1 和 a_2 为“光栏”边界点，p_1 与 a_1、p_1 与 a_2 的连线为以 p_1 为顶点的扇形的两条边，这就定义了一个扇形(这个扇形的口朝向曲线的前进方向，边长是任意的)。通过 p_1 并在扇形内的所有直线都具有这种性质，即 p_1p_2 上各点到这些直线的垂距都不大于 d/2。

(2)若 p_3 点在扇形内，则舍去 p_2 点。然后连接 p_1 和 p_3，过 p_3 作 p_1p_2 的垂线，该垂线与前面定义的扇形边交于 c_1 和 c_2。在垂线上找到 b_1 和 b_2 点，使 $p_3b_1=p_3b_2=d/2$，若 b_1 或 b_2 点((图 3—37 中为 b_2 点)落在原扇形外面，则用 c_1 或 c_2 取代(图 3—37 中由 c_2 取代 b_2)。此时用 p_1b_1 和 p_1c_2 定义一个新的扇形，这当然是口径(b_1c_2)缩小了的“光栏”。

(3)检查下一节点，若该点在新扇形内，则重复第(2)步；直到发现有一个节点在最新定义的扇形外为止。

(4)当发现在扇形外的节点，如图 7—20 中的 p_4，此时保留 p_3 点，以 p_3 作为新起点，重复(1)～(3)。如此继续下去，直到整个点列检测完为止。所有被保留的节点(含首、末点)，顺序地构成了简化后的新点列。

4. 几种方法的比较

如果某种矢量数据的压缩算法既能精确地表示数据，又能最大限度地淘汰不必要的点，那就是一种好的算法。具体可以依据简化后曲线的总长度、总面积、坐标平均值等与原始曲线的相应数据的对比来判别。

通过分析可以发现，大多数情况下道格拉斯——普克法的压缩算法较好，但必须在对整条曲线数字化完成后才能进行，且计算

量较大;光栏法的压缩算法也很好,并且可在数字化时实时处理,每次判断下一个数字化的点,且计算量较小;垂距法算法简单,速度快,但有时会将曲线的夹角去掉。

§7.5 空间数据的结构转换

1. 矢量—栅格转换

由于矢量数据的点到栅格数据的点只是简单的坐标变换,所以,这里主要介绍线和面(多边形)的矢量数据向栅格数据的转换。

(1)线的栅格化方法。线是由多个直线段组成的,因此,线的栅格化的核心就是直线段如何由矢量数据转换为栅格数据。

设直线段的两端点坐标转换到栅格数据的坐标系后为(x_A,y_A),(x_B,y_B)。则栅格化的两种常用方法为DDA法(数字微分分析法)和Bresenham法。

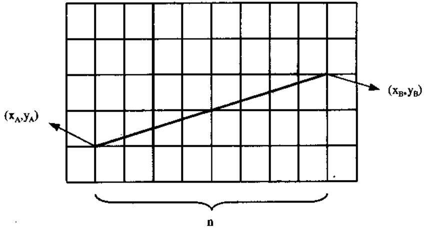


图7—21 DDA法示意图

——DDA法(数字微分分析法)。如图7—21所示,设(x_A,y_A),(x_B,y_B)与栅格网的交点为(x_i,y_i),则

$$\begin{cases} x_{i+1}=x_i+\dfrac{x_B-x_A}{n}=x_i+\Delta x \\ y_{i+1}=y_i+\dfrac{y_B-y_A}{n}=y_i+\Delta y \end{cases}$$

其中：

$$n=\max(|x_B-x_A|,|y_B-y_A|);$$

$$\Delta x=\frac{x_B-x_A}{n};\Delta y=\frac{y_B-y_A}{n};$$

$$x_0=x_A;y_0=y_A;x_n=x_B;y_n=y_B。$$

这样从 i＝0 计算到 i＝n－1，即可得直线与格网的 n 个交点坐标，对其取整就是该点的栅格数据了。

该方法的基本依据是直线的微分方程，即 dy/dx＝常数。其本质是用数值方法解微分方程，通过同时对 x 和 y 各增加一个小增量来计算下一步的 x，y 值，即这是一种增量算法。

在该算法中，必须以浮点数表示坐标，且每次都要舍入取整，因此，尽管算法正确，但速度不够快。

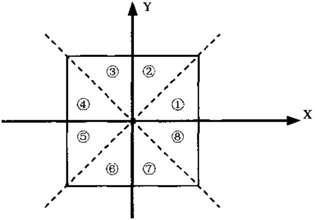


图 7—22　8 卦限示意图

——Bresenham 算法。该算法原来是为绘图机设计的，但同样适合于栅格化。该算法构思巧妙，只需根据由直线斜率构成的误差

项的符号，就可确定下一列坐标的递增值。

根据直线的斜率，把直线分为 8 个卦限(如图 7—22)。下面举斜率在第一卦限的情况为例，其余卦限的情况类似。

该算法的基本思路可描述为：如图 7—23，若直线的斜率为 $1/2 \leqslant \triangle y/\triangle x \leqslant 1$，则下一点取(1,1)点，若 $0 \leqslant \triangle y/\triangle x < 1/2$，则下一点取(1,0)点。

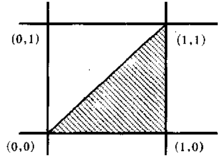


图 7—23 Bresenham 算法的思路

在算法实现时，令起始的误差项为 $e=-1/2$，然后在推断出下一点后，令 $e=e+\triangle y/\triangle x$，若 $e \geqslant 1/2$ 时，$e=e-1$。这样只要根据 e 的符号就可确定下一点的增量，即：若 $e \geqslant 0$，取(1,1)点；若 $e < 0$，取(1,0)点

例如，一直线的斜率为 1/3(如图 7—24)。

起始点：$e=-1/2$，　　　　取点①

第 2 点：$e=-1/2+1/3=-1/6<0$，　　　　取点②

第 3 点：$e=-1/6+1/3=1/6>0$，　　　　取点③

第 4 点：$e=1/6+1/3=1/2>0$，　　　　取点④

因 $e \geqslant 1/2$，所以，$e=1/2-1=-1/2$。

依次进行，直到到达直线的另一端点。

这种算法不仅速度快、效果好，而且可以证明。

2. 面(多边形)的栅格化方法

——内部点扩散法。由一个内部的种子点，向其 8 个方向的邻

点扩散。判断新加入的点是否在多边形边界上，如果是，不作为种子点，否则当作新的种子点，直到区域填满，无种子点为止。

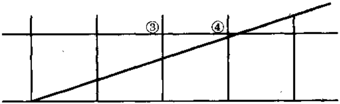


图 7—24 Bresenham 算法示例

该算法比较复杂，而且可能造成不连通(如图 7—25)，若多边形不完全闭合时，会扩散出去。

——扫描法。如图 7—26，按扫描线的顺序，计算多边形与扫

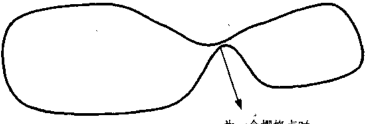


图 7—25 内部点扩散法的不连通情况

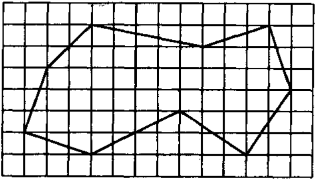


图 7—26 扫描法示意图

描线的相交区间，再用相应的属性值填充这些区间，即完成了多边形的栅格化。

这种算法的缺点是计算量较大。

——边填充算法。其基本思想是：对于每一条扫描线和每条多边形边上的交点，将该扫描线上交点右方的所有象素与多边形的属性值取补。对多边形的每条边作此处理，多边形的方向任意。图7—27是一个简单的例子。

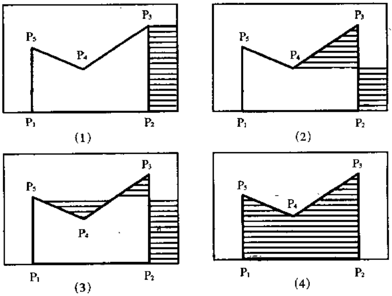


图7—27 边填充算法示意图

本算法的优点是算法简单，缺点是对于复杂图形，每一像素可能被访问多次，增加了运算量。为了减少边填充算法访问像素的次数，可引入栅栏。

所谓栅栏指的是一条与扫描线垂直的直线，栅栏位置通常取多边形的顶点，且把多边形分为左右两半。栅栏填充算法的基本思路是：对于每个扫描线与多边形的交点，将交点与栅栏之间的像素

用多边形的属性值取补。若交点位于栅栏左边，则将交点右边，栅栏左边的所有像素取补；若交点位于栅栏的右边，则将栅栏右边，交点左边的像素取补。图 7—28 是该算法的示意图。

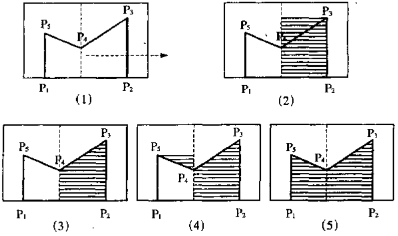


图 7—28 栅栏填充算法示意图

2. 栅格—矢量转换

栅格数据到矢量数据转换的一般过程可描述为：

——二值化。由于扫描后的图像是以不同灰度级存储的，为了进行栅格数据矢量化的转换，需压缩为两级（0 和 1），这就称为二值化。

二值化的关键是在灰度级的最大和最小值之间选取一个阈值，当灰度级小于阈值时，取值为 0，当灰度级大于阈值时，取值为 1。阈值可根据经验进行人工设定，虽然人工设定的值往往不是最佳阈值，但在扫描图比较清晰时，是行之有效的。当扫描图不清晰时，需由灰度级直方图来确定阈值，其方法为：

设 M 为灰度级数，P_k 为第 k 级的灰度的概率，n_k 为某一灰度级的出现次数，n 为像元总数，则有：

$$Pk = n_k / n \quad k = 1, \cdots, M$$

对于地图，通常在灰度级直方图上出现两个峰值（见图 7—29），这时，取波谷处的灰度级为阈值，二值化的效果较好。

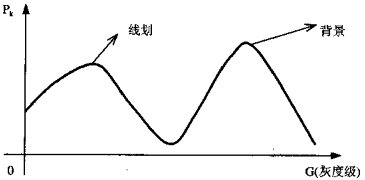


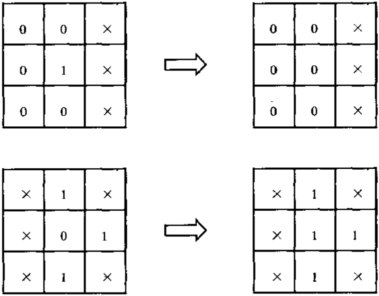
图 7—29　地图上的灰度组曲线

——二值图像的预处理。对于扫描输入的图幅，由于原稿不干净等原因，总是会出现一些飞白、污点、线划边缘凹凸不平等。除了依靠图像编辑功能进行人机交互处理外，还可以通过一些算法来进行处理。

例如用 3×3 的像素矩阵，规定各种情况的处理原则，图 7—30 是两个简单的例子：

除了上述方法外，还可用其他许多方法。例如，对于飞白和污点，给定其最小尺寸，不足的消除；对于断线，采取先加粗后减细的方法进行断线相连；用低通型滤波进行破碎地物的合并，用高通滤波提取区域范围等等。

——细化。所谓细化就是将二值图像像元阵列逐步剥除轮廓边缘的点，使之成为线划宽度只有一个像元的骨架图形。细化后的图形骨架既保留了原图形的绝大部分特征，又便于下一步的跟踪处理。



其中"×"表示任何像素值

图 7—30 二位图像的预处理

细化的基本过程是:(1)确定需细化的像元集合;(2)移去不是骨架的像元;(3)重复,直到仅剩骨架像元。

细化的算法很多,各有优缺点。经典的细化算法是通过 3×3 的像元组来确定如何细化的。其基本原理是,在 3×3 的像元组中,凡是去掉后不会影响原栅格影像拓扑连通性的像元都应该去掉,反之,则应保留。3×3 的像元共有 2^8 即 256 种情况,但经过旋转,去除相同情况,共有 51 种情况,其中只有一部分是可以将中心点剥去的,如图 7—31(1)、(2)是可剥去的,而(3)、(4)的中心点是不可剥去的。通过对每个像元点经过如此反复处理,最后可得到应保留的骨架像元。

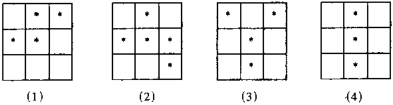


图 7—31　3×3 像元的分类情况

如果是对扫描后的地图图像进行细化处理，应符合下列基本要求：

· 保持原线划的连续性；

· 线宽只为一个像元；

· 细划后的骨架应是原线划的中心线；

· 保持图形的原有特征。

——追踪。细化后的二值图像形成了骨架图，追踪就是把骨架转换为矢量图形的坐标序列。其基本步骤为：

①从左向右，从上向下搜索线划起始点，并记下坐标。

②朝该点的 8 个方向追踪点，若没有，则本条线的追踪结束，转①进行下条线的追踪；否则记下坐标。

③把搜索点移到新取的点上，转②。

需注意的是，已追踪点应作标记，防止重复追踪。

——拓扑化。为了进行拓扑化，需找出线的端点和结点，以及孤立点。

①孤立点：8 邻城中没有为 1 的像元。如图 7—32(1)。

②端点：8 邻城中只有一个为 1 的像元。如图 7—32(2)。

③结点：8 邻城中有三个或三个以上为 1 的像元。如图 7—32(3)。

在追踪时加上这些信息后，就可形成结点和孤段，就可用矢量数据的自动拓扑方法进行拓扑化了。

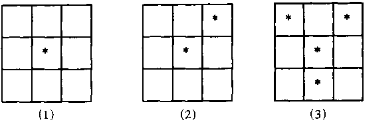


图 7—32 孤立点、端点和结点示意图

§7.6 空间数据的内插

1. 空间数据的插值

用各种方法采集的空间数据往往是按用户自己的要求获取的采样观测值，亦既数据集合是由感兴趣的区域内的随机点或规则网点上的观测值组成的。但有时用户却需要获取未观测点上的数据，而已观测点上的数据的空间分布使我们有可能从已知点的数据推算出未知点的数据值。

在已观测点的区域内估算未观测点的数据的过程称为内插；在已观测点的区域外估算未观测点的数据的过程称为外推。

空间数据的内插和外推在 GIS 中使用十分普遍。一般情况下，空间位置越靠近的点越有可能获得与实际值相似的数据，而空间位置越远的点则获得与实际值相似的数据的可能性越小。下面介绍一些常用的内插方法。

(1)边界内插。使用边界内插法时，首先要假定任何重要的变化都发生在区域的边界上，边界内的变化则是均匀的、同质的。

边界内插的方法之一是泰森多边形法。泰森多边形法的基本原理是，未知点的最佳值由最邻近的观测值产生。如图 7—33 所示，A、B、C、D 四点分别落在各自的虚线构成的多边形中，则这些多边形中每一点的数据值都相同，分别为 A、B、C、D 四点的数据值。

泰森多边形的生成算法见§8.6。

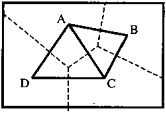


图 7—33　泰森多边形的基本原理

(2)趋势面分析。趋势面分析是一种多项式回归分析技术。多项式回归的基本思想是用多项式表示线或面，按最小二乘法原理对数据点进行拟合，拟合时假定数据点的空间坐标的 X、Y 为独立变量，而表示特征值的 Z 坐标为因变量。

当数据为一维时，可用回归线近似表示为：

$$Z=a_0+a_1X \tag{7-9}$$

其中，a_0、a_1 为多项式的系数。当 n 个采样点方差和为最小时，则认为线性回归方程与被拟合曲线达到了最佳配准，如图 7—34 所示，即：

$$\sum_{i=1}^{n}(\hat{Z}_i-Z_i)^2=\min \tag{7-10}$$

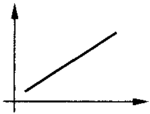


图 7—34　最佳配准情况



图 7--35　复杂情况

当数据以更为复杂的方式变化时，如图 7—35 所示。在这种情况下，需要用到二次或高次多项式：

$$Z=a_0+a_1X+a_2X^2 \quad \text{（二次曲线）} \tag{7-11}$$

在 GIS 中，数据往往是二维的，在这种情况下，需要用到二元二次或高次多项式：

$$Z=a_0+a_1X+a_2Y+a_3X^2+a_4XY+a_5Y^2 \quad (二次曲面) \quad (7—12)$$

多项式的次数并非越高越好，超过 3 次的多元多项式往往会导致奇异解，因此，通常使用二次多项式。

趋势面是一种平滑函数，难以正好通过原始数据点，除非数据点数和多项式的系数的个数正好相同。这就是说，多重回归中的残差属正常分布的独立误差，而且趋势面拟合产生的偏差几乎都具有一定程度的空间非相关性。

(3)局部内插。在 GIS 中，实际的连续空间表面很难用一种数学多项式来描述，因此，往往使用局部内插技术，既利用局部范围内的已知采样点的数据内插出未知点的数据。常用的有线性内插、双线性多项式内插、双三次多项式(样条函数)内插。

——线性内插。线性内插的多项式函数为：

$$Z=a_0+a_1X+a_2Y \quad (7—13)$$

只要将内插点周围的 3 个数据点的数据值带入多项式，既可解算出系数 a_0、a_1、a_2。

——双线性多项式内插。双线性多项式内插的多项式函数为：

$$Z=a_0+a_1X+a_2Y+a_3XY \quad (7—14)$$

只要将内插点周围的 4 个数据点的数据值带入多项式，既可解算出系数 a_0、a_1、a_2、a_3。

如果数据是按正方形格网点布置的(如图 7—36)，则可用简单的公式既可计算出内插点的数据值。

设正方形的四个角点为 A、B、C、D，其相应的特征值为 Z_A、Z_B、Z_C、Z_D，P 点相对于 A 点的坐标为 d_x、d_y，则插值点的特征值 Z 为：

$$Z=(1-\frac{d_x}{L})\cdot(1-\frac{d_y}{L})\cdot Z_A+(1-\frac{d_y}{L})\cdot\frac{d_x}{L}\cdot Z_B+\frac{d_x}{L}\cdot\frac{d_x}{L}\cdot Z_C+(1-\frac{d_x}{L})\cdot\frac{d_x}{L}\cdot Z_D \quad (7—15)$$

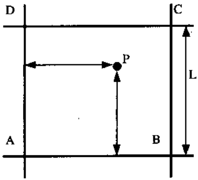


图 7—36　采样点成格网分布的情况

——双三次多项式(样条函数)内插。双三次多项式是一种样条函数。样条函数是一种分段函数,对于 n 次多项式,在边界处其 n－1 阶导数连续。因此,样条函数每次只用少量的数据点,故内插速度很快;样条函数通过所有的数据点,故可用于精确的内插,可以保留微地貌特征;样条函数的 n－1 阶导数连续,故可用于平滑处理。

双三次多项式内插的多项式函数为:

$$Z=a_0+a_1X+a_2Y+a_3X^2+a_4XY+a_5Y^2+a_6X^3+a_7XY^2$$
$$+a_8X^2Y^2+a_9X^2Y+a_{10}Y^3+a_{11}XY^3$$
$$+a_{12}X^3Y+a_{13}X^2Y^3+a_{14}X^3Y^2+a_{15}X^3Y^3 \quad (7—16)$$

将内插点周围的 16 个点的数据带入多项式,可计算出所有的系数。

(4)移动平均法。在未知点 X 处内插变量 Z 的值时,最常用的方法之一是在局部范围(或称窗口)内计算多个数据点的平均值。即:

$$\hat{Z}(X)=\frac{1}{n}\sum_{i}^{n}Z(X_i) \quad (7—17)$$

对于二维平面的移动平均法也可用相同的公式,但位置 X_i 应被坐标矢量 X_i 代替。

窗口的大小对内插的结果有决定性的影响。小窗口将增强近距离数据的影响；大窗口将增强远距离数据的影响，减小近距离数据的影响。

当观测点的相互位置越近，其数据的相似性越强；当观测点的相互位置越远，其数据的相似性越低。因此，在应用移动平均法时，根据采样点到内插点的距离加权计算是很自然的。这就是加权移动平均法，即：

$$\hat{Z}(X)=\frac{1}{n}\sum_{i}^{n}\lambda_i Z(X_i) \qquad (7-18)$$

其中，λ_i 是采样点 i 对应的权值，常取的形式有：

$$\lambda_i=\frac{1}{d_i^2}$$

$$\lambda_i=\left(\frac{R-d_i}{d_i}\right)^2 \qquad (7-19)$$

$$\lambda_i=e^{-d_i^2/R^2}$$

其中 R 为彩样点到内插点的最大距离。

加权平均内插的结果随使用的函数及其参数、采样点的分布、窗口的大小等的不同而变化。通常使用的采样点数为 6—8 点。对于不规则分布的采样点需要不断地改变窗口的大小、形状和方向，以获取一定数量的采样点。

2. 格网 DEM 的生成

格网 DEM 的数据可直接从解析测图仪获取，因而下面介绍的是如何由离散点来构建格网 DEM 的方法。

(1)离散点构建格网 DEM 的基本思路。离散点构格网 DEM 是在原始数据呈离散分布，或原有的格网 DEM 密度不够时需使用的方法。其基本思路是：选择一合理的数学模型，利用已知点上的信息求出函数的待定系数，然后求算规则格网点上的高程值。

(2)离散点构建格网 DEM 的内插算法。离散点构格网 DEM 所采用的是内插算法，插值的方法很多，如按距离加权法、多项式

内插法、样条函数内插法、多面函数法等等。大量的实验证明，由于实际地形的非平稳性，不同的内插方法对 DEM 的精度并无显著影响，主要取决于原始采样点的密度和分布。下面介绍简单而常用的线性内插法和双线性多项式内插法。

线性内插的数学模型为：

$$Z=a_0+a_1x+a_2y \quad (7—20)$$

双线性多项式内插的数学模型为：

$$Z=a_0+a_1x+a_2y+a_3xy \quad (7—21)$$

其中，x，y 为平面坐标，Z 为高程，a_0、a_1、a_2、a_3 为待定系数。

只要将与插值点距离最近的三个点（对线性内插）或四个点（对双线性多项式内插）的坐标值和高程值代入方程，即可解出全部系数，然后用插值点的坐标带入方程，即可计算出该点的高程值。

距离加权法的数学模型为：

$$Z=\sum_{i=1}^{n}P_i \cdot Z_i / \sum_{i=1}^{n}P_i \quad (7—22)$$

为了计算点（x，y）的高程 Z，可取该点周围的 n 个点（i=1，…，n），按上述公式计算。Z_i 为第 i 点的高程，P_i 为第 i 点的权值，计算公式为：

$$P_i=1/d_i^u \quad (i=1,2,\cdots,n) \quad (7—23)$$

其中 u 是一个大于 0 的正数，通常取 1 或 2，d_i 是 i 点（x_i，y_i）到格网点（x，y）的距离，即：

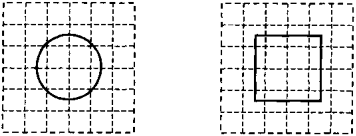
$$d_i=\sqrt{(x_i-x)^2+(y_i-y)^2} \quad (i=1,2,\cdots,n) \quad (7—24)$$

（3）离散点构建格网 DEM 的内插算法中原始数据点的选取。

在构建格网 DEM 的内插方法中，经常需要选取与插值点距离最近的若干个点。如图 7—38（1），如果通过计算各离散点与待插值点的距离，然后选取距离最短的若干个点的方法来选取，尽管方法正确，但计算量大，影响插值的速度。因此，可以在插值点上建

立一正方形的选取框，如图 7—38(2)，这样通过简单的坐标值比较就可找出落入框内的数据点。当落入框内的数据点较多时，可缩小框的尺寸；反之，增大框的尺寸。选取框尺寸的初始值可根据图幅中原始数据点的密度来确定。设图幅的面积为 A，共有 N 个数据点，则每点的平均面积 A_0 为：

$$A_0 = A/N \qquad (7—25)$$



(1)圆形选取框　　　(2)矩形选取框

图 7—38

若需要选取插值点附近的 K 个点，则选取框的面积应为：

$A_1 = K \cdot A_0$。

$A_1^{1/2}$就是该选取框的边长。在实际运算时，边长可再大一些，以尽量保证落入选取框中的数据点数大于或等于 K，这样在缩小选取框时，只要对原框内数据点再判断即可。

(4)由矢量等高线数据构建格网 DEM 时需注意的问题。由矢量等高线数据构建格网 DEM 时，通常把矢量等高线数据当作离散点数据，然后用上述离散点构建格网 DEM 的方法计算。出现的主要问题是应该水平的区域(如海洋、湖泊、平地等)往往出现不应该的不规则起伏，其原因是在计算水平区域的 DEM 格网点高程时，所使用的原始数据点不合理(如在计算湖泊内的 DEM 点时，使用了周围陆地的高程点进行内插)。解决的方法通常有以下几种：

—— 在构建格网 DEM 之前，获取构成水平区域的多边形，及

其高程值。在构建格网 DEM 时，如果格网点落在这些多边形中，就直接取相应多边形的高程作为该格网点的高程。这样就可保证所构建的格网 DEM 在水平区域是水平的。

——先使用普通的离散点构建格网 DEM 的方法计算，在生成格网 DEM 后，对水平区域作统一处理，即把落在水平区域多边形内的 DEM 格网点的高程，用该水平区域的高程代替。这样也可保证所构建的格网 DEM 在水平区域是水平的。

——通过控制用于 DEM 格网点高程内插的原始数据点，使得参加水平区域的 DEM 格网点计算的原始数据点的高程为水平区域本身的高程，这样内插出的高程值就等于水平区域本身的高程。

例如，在计算某一格网点的高程时，不以距离取参加计算的原始数据点，而是通过该格网点向 8 方向(间隔 45°)作射线，取射线与离该格网点最近的等高线的交点作为参加计算的原始数据点。由于包围水平区域的等高线必然具有相同的高程，所以计算水平区域格网点的高程时，所参加计算的原始数据点的高程都是一样的，这样可保证所构建的格网 DEM 在水平区域是水平的。这种方法可以称为射线法，它要求等高线在图幅内不能随意断开(图廓边除外)。

§7.7 图像数据的处理

1. 图像增强

一幅图像经过生成、复制、扫描、传输、变换后，由于多种因素的影响，图像的质量不能满足要求，这时就需要进行图像增强处理。图像增强的目的是改善图像的效果，以更适应人眼的观察或计算机的处理。

但是，对图像质量的评价并没有统一的标准，图像增强的方法往往带有一定的针对性。下面介绍一些基本的图像增强方法。

（1）灰度级的修整。直接修改图像像素点灰度级是一种简单而有效的图像增强方法，主要有两种形式：一种是灰度级校正，通过修改像素点的灰度级来补偿记录图像时的不均匀曝光；另一种是灰度级变换，用统一的方法改变图像的灰度，以提高图像的质量。

①灰度级校正。图像记录系统应将物体的亮度单调地映射成图像的灰度级。在理想情况下，上述映射关系不随像素点的位置而变化，然而，实际上却随像素点的位置而变化，即是不均匀曝光。

假定均匀曝光下图像的灰度级为 f(x,y)，而实际上非均匀曝光下图像的灰度级为：

$$g(x,y)=e(x,y)\cdot f(x,y) \qquad (7—26)$$

其中 e(x,y)描述了曝光的非均匀性。

为了确定 e(x,y)，可使用一个已知亮度的均匀场面的图像来核准图像记录系统。设这个均匀场面经过均匀曝光后的灰度级为常数 C，而这个均匀场面经过非均匀曝光后的图像的灰度级为 gc(x,y)，则：

$$e(x,y)=gc(x,y)/C \qquad (7—27)$$

这样根据 e(x,y)就能校正该系统得到的任何图像。

②灰度变换。当图像成像时曝光不足或过度，图像记录设备的范围太窄等因素，都会产生对比不足的问题，使图像的细节分辨不清。为此需对每一像素的灰度级进行变换，扩大图像灰度的范围，达到图像增强的目的。

设原图像中像素点(x,y)处的灰度级为 f(x,y)，通过映射函数 T，生成的图像的灰度级为 g(x,y)，即：

$$g(x,y)=T[f(x,y)] \qquad (7—28)$$

· 线性灰度变换。将对比度较差的图像的灰度线性扩展，常能显著改善图像的质量。假定原图像 f(x,y)的灰度范围为[a,b]，变换后的图像 g(x,y)的灰度范围为[c,d]，则有：

$$g(x,y)=\frac{(d-c)[f(x,y)-a]}{b-a}+c \qquad (7—29)$$

·非线性灰度变换。当用某些非线性函数，如对数、指数函数作为映射函数时，可实现图像灰度的非线性变换。对数变换的一般公式为：

$$g(x,y)=a+\frac{\ln[f(x,y)+1]/b}{\ln c} \qquad (7-30)$$

a、b、c 为可调参数。当希望对图像的低灰度区有较大的扩展，而对高灰度区进行压缩时，可采用此变换。

指数变换的形式为：

$$g(x,y)=b^{c[f(x,y)-a]}-1 \qquad (7-31)$$

这种变换能对图像的高灰度区以较大的扩展。

·灰度分层切片。此变换将图像中的某一段灰度范围抽取出来，转换成最大的灰度值。

$$g(x,y)=\begin{cases} MAX & a<f(x,y)<b \\ 0 & \text{其他} \end{cases} \qquad (7-32)$$

③直方图变换。直方图表示数字图像中每一灰度级与其出现的频数(具有该灰度级的像素的数目)间的统计关系，横坐标表示灰度级，纵坐标表示频数(或相对频数=频数/总像素数)。直方图能给出图像的概貌性描述，如图像的灰度范围、灰度级的大致分布情况等。

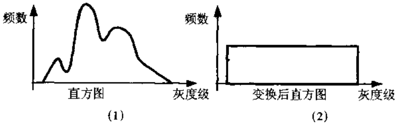


图 7—39 直方图均衡化

·直方图均衡化。如图 7—39(1)、(2)，把原图像的直方图变换为各灰度值频率固定的直方图称为直方图均衡化。

·直方图正态化。如果灰度的频率分布接近正态分布的形状，

通常认为适合于人眼观察，如图 7—40(1)、(2)所示。但如果把与正态分布形状相差较大的图像勉强进行直方图正态化时，往往会产生问题。如当原图像的某一灰度的频率很高，而正态分布所对应的该灰度值的频率变得较低，就会造成这部分信息被压缩和丢失的情况。因此，直方图正态化对于卫星图像那种原图像的动态范围窄，且不够鲜明的图像是非常有效的。

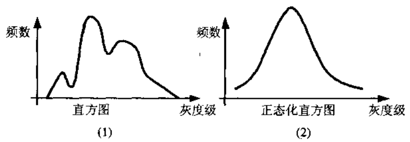


图 7-40　直方图正态化

(2)空域处理。

①平滑。图像在生成和传输过程中会受到各种噪声源的干扰和影响，使图像质量变差。反映在图像上，噪声使原本均匀和连续变化的灰度突然变大或变小，形成一些虚假的物体边缘或轮廓。抑制或消除这些噪声而改善图像质量的过程称为图像的平滑。

・邻域平均法。在邻域平均法中，假定图像是由许多灰度恒定的小块组成，相邻像素间有很强的空间相关性，而噪声是统计独立地加到图像上的。因此，可用像素邻域内多个像素灰度值的平均值来代表原来的灰度值。根据求平均方式的不同，主要有以下几种形式：

简单平均法。设图像中某像素的灰度值为 f(x,y)，它的邻域 S 为 M×N 的矩形窗口，则平滑后该点的灰度值为：

$$f'(x,y)=\frac{1}{M\cdot N}\sum_{(u,v)\in S}f(u,v) \qquad (7—33)$$

阈值平均法。设图像中某像素的灰度值为 f(x,y)，它的邻域

为 M×N 的矩形窗口，则平滑后该点的灰度值为：

$$g(x,y)=\begin{cases}f'(x,y) & 若|f(x,y)-f'(x,y)|>T \\ f(x,y) & 其他\end{cases} \quad (7—34)$$

T 为给定的阈值。f'(x,y)的计算方法同简单平均法。

K—近邻平均法。在一个 M×N 的窗口中，属于同一个图像的像素的灰度值会高度相关。因此，窗口中心像素的灰度值可用窗口内与中心点灰度最接近的 K 个邻点的平均灰度来代替。一般而言，K 值越小，噪声方差降低越小，细节保持较好；而较大的 K 值平滑噪声效果好，但会使图像模糊。

· 低通滤波法。从频谱上看，噪声特别是随机噪声是一种具有较高频率分量的信号。平滑的目的就是通过一定的手段滤去这类信号。一个很自然的想法就是使图像经过一个二维的低通数字滤波器，让高频信号得到较大的衰减。在空间域上进行的这种滤波实际上就是对图像和滤波器的冲击响应函数进行卷积。

设图像为 f(x,y)，滤波器的冲击响应函数为 H(x,y)，则卷积表达式为：

$$g(u,v)=\sum_{x}\sum_{y}f(x,y)H(u-x+1,v-y+1) \quad (7—35)$$

· 中值滤波法。中值滤波的思想是对一个窗口内的所有像素的灰度值进行排序，取排序结果的中间值作为原窗口中心点处像素的灰度值。这种平滑方法对脉冲干扰和校验类干扰噪声的效果较好。

中值滤波的关键在于选择合适的窗口大小和形状。但一般很难事先确定窗口的尺寸，通常是从小到大进行多次尝试。窗口的形状可选为正方形，也可选为十字形。

②尖锐化。在图像判断和识别中，需要有边缘鲜明的图像。图像尖锐化技术常用来对图像的边缘进行增强。

· 微分法。在图像的判断和识别中，边缘是由不同灰度级的相邻像素点构成的。因此，若想增强边缘，就应该突出相邻点间的灰

度级变化。微分运算可用来求信号的变化率，具有加强高频分量的作用。如果将其应用在图像上，可使图像的轮廓清晰。由于常常无法事先确定轮廓的取向，因而在挑选用于轮廓增强的微分算子时，必须选择那些不具备空间方向性和具有旋转不变性的线性微分算子。

·高通滤波法。由于边缘是由灰度级跳变点构成的，因此，具有较高的空间频率。所以可用高通滤波法让高频分量顺利通过，使低频分量得到抑制。通过增强高频分量，使图像的边缘变得清晰，实现图像的尖锐化。

这一思想反应在空间域的处理中就是让图像和高通滤波器的冲击响应函数进行卷积。所用的表达式与低通滤波法中所用的相同，只是冲击响应函数不同。

(3)频域处理。频域处理是指根据一定的图像模型，对图像的傅立叶频谱的各个频段进行不同程度的修改的技术。通常总是假设：

引起图像质量下降的噪声在图像的傅立叶频谱中占据的是高频段；

图像的边缘在傅立叶频谱中占据的也是高频段。

图像的主体或图像中灰度变化较缓的区域在频谱中占据的是低频段。

①频域中的平滑。平滑的目的是滤去噪声，即保留低频段，使高频段受到大的抑制。最常用的方法是低通滤波，其数学表达式为：

$$G(u,v)=H(u,v)\cdot F(u,v) \qquad (7-36)$$

其中，F(u,v)是图像的傅立叶频谱，H(u,v)是低通滤波的转移函数(即频谱响应)。对低通滤波而言，H(u,v)应使高频抑制，使低频通过。常用的H(u,v)函数如：理想低通滤波器、Butterworth滤波器、指数性滤波器、梯形滤波器等，具体的数学表达式请参阅有关图像处理书籍。

②频域中的锐化。锐化的目的是突出边缘，即保留高频段，而使低频段受到大的抑制。最常用的方法是高通滤波，其数学表达式为：

$$G(u,v)=H(u,v)\cdot F(u,v)$$

其中，F(u,v)是图像的傅立叶频谱，H(u,v)是高通滤波的转移函数（即频谱响应）。对高通滤波而言，H(u,v)应使低频抑制，而使高频通过。常用的H(u,v)函数如：理想高通滤波器、Butterworth滤波器、指数性滤波器、梯形滤波器等，具体的数学表达式请参阅有关图像处理书籍。

③同态滤波。同态滤波的目的是通过对图像作非线性变换，使构成图像的非可加性因素成为可加性的，从而容易进行滤波处理。下面举一例子加以说明。

设图像f(x,y)由照射分量I(x,y)和反射分量r(x,y)的乘积构成，即：

$$f(x,y)=I(x,y)\cdot r(x,y) \qquad (7—37)$$

因而有：$\ln[f(x,y)]=ln[I(x,y)]+ln[r(x,y)]$

I(x,y)描述的是照射源的特性，一般假设是缓变的；而r(x,y)描述的是景物的特性，随物体的细节在空间上作快速变化。

如果对ln[f(x,y)]作傅立叶变换，则其一部分是低频段的ln[I(x,y)]频谱，另一部分是高频段的ln[r(x,y)]频谱。可以用同一个滤波器进行滤波处理，而达到图像增强的目的。

设滤波器的频谱响应为H(u,v)，其低频特性可根据对ln[I(x,y)]的增强确定，其高频特性可根据对ln[r(x,y)]的增强确定。经过滤波处理后，再将ln[f(x,y)]进行反变换，就可得到增强后的f(x,y)。

(4)伪彩色增强。人眼对灰度级是极不敏感的，通常可分辨十几到二十几个灰度级，但却可以分辨出数千种的彩色。因此，可以用彩色来增强灰度图像。

伪彩色技术不是观察物体的真正的颜色，而是根据图像的灰

度级，通过一种关系来指定相应的颜色，和物体本身的颜色毫无关系。

①灰度级到彩色的处理。根据图像像素的灰度级，建立三个独立的变换关系，将灰度级变换为红、绿、蓝色调。其数学表达式为：

$$(r,g,b)=(R[f(x,y)],\ G[f(x,y)],B[f(x,y)])$$

②滤波方法。这种彩色处理技术的目的是根据频率的成分对一幅图像的各个区域进行彩色编码。具体步骤是：先利用三种不同的滤波器分别得到三个频率范围内的频率分量；然后对上述三种频率分量分别作傅立叶变换，得到变换后的三幅图像；再将这三幅图像分别作为显示系统的红色、绿色、蓝色的输入端。有时在送入输入端之前，还进行一些附加处理，如直方图均衡化。

通常使用低通、带通、高通滤波获得三个范围的频率分量。带通滤波使在一定范围内的频率通过，而使在范围外的频率得到抑制或衰减。

2. 二值图像处理

在数字图像中，二值图像占有非常重要的地位。在对地图、文字的扫描数字化和识别时，通常都是当作二值图像来处理的。在二值图像中，通常 1 表示图形，0 表示背景。

(1)图像的二值化。二值化是为了从图像中分离出对象物，即把图形和背景作为二值图像对待。图像二值化可以用下列阈值处理方法进行(如图 7—41)。

$$f_t(i,j)=\begin{cases}1 & 当\ f(i,j)\geq t\ 时\\ 0 & 当\ f(i,j)<t\ 时\end{cases} \tag{7—38}$$

这里的问题是阈值 t 的确定方法。下面介绍主要的方法。

①状态法。求出给定图像的灰度值直方图，在具有两个峰值(对应于图形和背景)的情况下，可以在峰之间的谷底确定 t 值。如图 7—42 。

②微分直方图法。这是设想图像的图形和背景的边界处的灰度值急剧变化。这样就可不直接利用图像的灰度值，而是利用微分

值(灰度的变化率)来确定阈值。如图 7—43 。

设图像中某一像素的值为S,计算这个像素的微分值,如取邻域各像素与该像素灰度值之差的最大值,或取各个灰度值之差的绝对值之和;求出图像中具有灰度值S的所有像素的微分值之和。对所有的像素作同样的处理,即得到微分直方图。

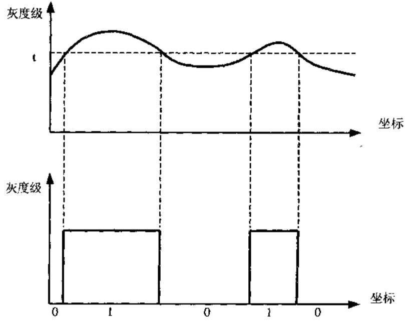


图 7--41 阈值处理

如果选择微分直方图中最高值的灰度值,就可认为这一灰度值对应于灰度变化率最高的部分。这一方法在图形的边界处于一定的灰度值范围时,是比较有效的。但往往边界附近灰度值的变化比较复杂,这时就并不很有效。

③可变阈值法。当由于图像位置不同而平均灰度值不同时,用单一的阈值不能有效地对整个图像进行二值化。这时,可让阈值随着图像的不同部分而变化,从而对整个图像进行有效处理。

(2)二值图像的平滑去噪。对于扫描输入或经过二值化的地图图像，由于原稿不干净或图像质量较差，总是伴随着毛刺、污点、线划边缘凸凹不平等，通过二值图像的平滑去噪，可以去掉孤立的毛刺、黑斑，平滑线划的边缘，填补面状目标内的小孔等，提高图像的质量。

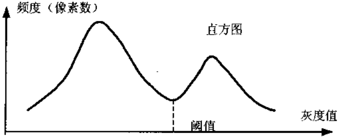


图 7—42　状态法确定阈值

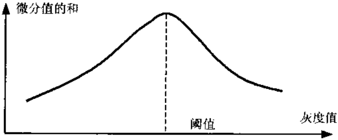


图 7—43　微分值方图法确定阈值

一般的平滑处理方法是，采用n×n的辅助矩阵(n一般为3～5)为模板，逐行、逐列与二值图像匹配，根据辅助矩阵中0、1像元的分布，使处于矩阵中心的像元从“0”变成“1”，或从“1”变为“0”。

①去毛刺。通常采用图7—44所示的3×3的矩阵，包括其3次90°旋转形成的矩阵。当矩阵模板在二值图像上移动时，只要二值图像与模板匹配，就把模板中心的像元从“1”变为“0”。

0	0	0
0	1	0
×	×	×

图 7-44 去毛刺的模板

②线划平滑与孔洞填补。通常采用图 7—45 所示的 3×3 的矩阵，包括其 3 次 90°旋转形成的矩阵。当矩阵模板在二值图像上移动时，只要二值图像与模板匹配，就把模板中心的像元从"0"变为"1"。

×	1	×
1	0	1
×	×	×

图 7-45 线划平滑与孔洞填补的模板

③去除独立污点。独立污点(小黑斑)是指其周围的像素都为"0"，而其本身的大小又可被一个 n×n 的矩阵所覆盖。则建立 n×n 的矩阵，使其在图像上移动。移动时计算矩阵周边像素值的和，若其和为 0，则将矩阵中心 n－2×n－2 区域的像素值赋为"0"。

(3)二值图像的边缘特征提取。二值图像的边缘特征提取是 GIS 图像处理中常用的技术，例如可以用于对地图扫描数字化后的数据进行面状要素的提取。二值图像边缘特征提取的过程实际上是寻找像素灰度值急剧变化的位置的过程，并在这些位置上将像素值置为"1"，其余位置上的像素值置为"0"，从而求出目标的边界线。

二值图像的边缘特征提取是用数学算子实现的，如 Sobel、Prewitt、Kirsch、拉普拉斯等多种算子。这些算子都是以一个 3×3 的模板与图像中 3×3 的区域相乘，得到的结果作为图像中这个区域中心位置的边缘强度。在计算出图像中每一个像素的边缘强度后，将边缘强度大于一定值的点提取出来，并赋以像素值"1"，其余

赋以像素值“0”。

设 f(i,j)是(i,j)处的像素值,(i,j)位置处的边缘强度通常用差分值或其函数来表示。简单的差分算法有:

x 方向差分值:$\triangle xf(i,j)=f(i,j)-f(i,j-1)$

y 方向差分值:$\triangle yf(i,j)=f(i,j)-f(i-1,j)$

边缘强度$=|\triangle xf(i,j)|+|\triangle yf(i,j)|$ 或

$=\triangle x^2 f(i,j)+\triangle y^2 f(i,j)$,等等。

各种算子的区别实际上只是计算差分的方法不同。现将几种常用的算子介绍如下:

· Prewitt 算子

(1)

1	0	−1
1	0	−1
1	0	−1

(2)

1	1	1
0	0	0
−1	−1	−1

图 7-46 Prewitt 算子的模板

计算△xf 的模板为图 7—46(1),计算△yf 的模板为图 7—46(2)。其特点是:对称于中心位置处各像素的权值相同。

△xf 、△yf 分别等于相应模板与图像中对应区域元素相乘之和。

即:

$\triangle xf=[f(i-1,j-1)-f(i-1,j+1)]+[f(i,j-1)-f(i,j+1)]+[f(i+1,j-1)-f(i+1,j+1)]$

$\triangle yf=[f(i-1,j-1)-f(i+1,j-1)]+[f(i-1,j)-f(i+1,j)]+[f(i-1,j+1)-f(i+1,j+1)]$

边缘强度$=|\triangle xf|+|\triangle yf|$或$=max(|\triangle xf|,|\triangle yf|)$。

· Sobel 算子。计算△xf 的模板为图 7—47(1),计算△yf 的模板为图 7—47(2)。其特点是:对称于中心位置处,左右或上下方

向上与对角线方向上各像素的权值不一样。

1	0	−1
2	0	−2
1	0	−1

(1)

1	2	1
0	0	0
−1	−2	−1

(2)

图 7—47 Sobel 算子的模板

· 拉普拉斯算子。计算$\triangle xf$和$\triangle yf$的模板为图 7—48。边缘强度$\triangle^2 f$等于模板与图像中对应区域元素相乘之和，是一个二次差分，具有各向相同性，即不依赖于边缘线的方向。

0	1	0
1	-4	1
0	1	0

图 7- 48 拉普拉斯算子的模板

· Kirsch 算子

5	5	5
−3	0	−3
−3	−3	−3

M_0

5	5	−3
5	0	−3
−3	−3	−3

M_1

5	−3	−3
5	0	−3
5	−3	−3

M_2

−3	−3	−3
5	0	−3
5	5	−3

M_3

$$
\begin{array}{|c|c|c|}\hline -3 & -3 & -3 \\ \hline -3 & 0 & -3 \\ \hline 5 & 5 & 5 \\ \hline \end{array}
\qquad
\begin{array}{|c|c|c|}\hline -3 & -3 & -3 \\ \hline -3 & 0 & 5 \\ \hline -3 & 5 & 5 \\ \hline \end{array}
$$

M_4　　　　M_5

$$
\begin{array}{|c|c|c|}\hline -3 & -3 & 5 \\ \hline -3 & 0 & 5 \\ \hline -3 & -3 & 5 \\ \hline \end{array}
\qquad
\begin{array}{|c|c|c|}\hline -3 & 5 & 5 \\ \hline -3 & 0 & 5 \\ \hline -3 & -3 & -3 \\ \hline \end{array}
$$

M_6　　　　M_7

图 7-49　Kirsch 算子的模板

这是一种最佳适配的边缘检测法。具有如图 7-49 所示的 0°、45°、90°、135°、180°、225°、270°、315°8 个方向上的模板。用 $M_0 \sim M_7$ 分别与图像的各对应元素相乘，取计算结果的最大值作为中央像素的边缘强度。

(4) 二值图像的细化。为了从二值图像中提取线状目标的坐标信息，需要将粗线转换成单一像元连接而成的细线，即进行二值图像的细化，二值图像的细化在图像处理中有着广泛的应用。

通常细化后的图形应具有如下性质：

- 线宽为 1 个像元。
- 细化后的线应尽量在原图形的中心。
- 应保持图形的连通性，不能因细化而断开。
- 细化后的线的缩小必须有一定的限度。
- 不因边界上的小凸凹而产生毛刺。
- 图像旋转后再处理，中心线的形状变化不大。
- 图形交叉部分中心线不畸变。

细化算法有多种多样，但对算法的评价却需要根据面对的具体问题来分析。有的算法速度快，但易产生断线；有的速度慢，不易产生断线，但却会产生毛刺；等等。这里只举一种算法。该算法基

于模板匹配思想，即用模板逐行逐列与图像配准。若与模板配准，则为骨架像素(粗线目标的中心线像素)，否则为轮廓像素(粗线目标的边缘像素)。

该算法有两个 3×3 模板，对于中心点 P，其模板如图 7-50 所示。

A	A	A
0	P	0
B	B	B

模板 1

A	A	A
A	P	0
A	0	S

模板 2

图 7-50　细化的模板

模板 1 还包括一次旋转 90°后的模板；模板 2 还包括一次、两次、三次旋转 90°后的模板。在模板中，A 或 B 中至少有一个像元为非 0 时，P 即为骨架像素，S 为已判定的骨架像素。

第八章　空间查询与空间分析

§8.1　空间数据的查询

1. 空间数据查询的含义

查询属于数据库的范畴，一般定义为作用在库体上的函数，返回满足条件的内容。查询是用户与数据库交流的途径。

查询是 GIS 用户最经常使用的功能，GIS 用户提出的很大一部分问题都可以以查询的方式解决，查询的方法和查询的范围在很大程度上决定了 GIS 的应用程度和应用水平。

目前 GIS 的空间查询主要有下列 4 种方法：

- 扩展关系数据库的查询语言(SQL)；
- 可视化空间查询；
- 超文本查询；
- 自然语言空间查询。

2. 扩展关系数据库的查询语言(SQL)

由于关系数据库具有严谨的数学基础和简洁的概念，在一般的事务性数据库中占有绝对的统治地位。在关系数据库中，几乎所有的功能都由查询语言(SQL)实现，关系数据库的查询语言(SQL)作为一种工业标准被广泛使用。

目前 GIS 的地理数据库大多是以传统的关系数据库为基础的，但 GIS 中的地理数据库是一种特殊的数据库，其最大的不同是具有"空间"概念。空间数据查询不仅应能方便地选取用户所需要的空间数据，而且要能以用户易于理解的形式把查询结果显示

出来。空间数据的查询语言作为用户与 GIS 交互的手段，决定了用户与 GIS 相互理解的程度。

SQL 语句通常是由关系运算组合而成的，非常适合于关系表的查询与操作，但并不支持空间运算。由于标准的 SQL 不支持空间概念，因此，不能进行空间数据的查询。

目前的空间数据查询语言是通过对标准 SQL 的扩展来形成的，即在数据库查询语言上加入空间关系查询。为此需要增加空间数据类型(如点、线、面等)和空间操作算子(如求长度、面积、叠加等)。在给定查询条件时也需含有空间概念，如距离、邻近、叠加等。例如，“显示与价值超过 6 万的地块相交的土壤图”，可表示为：

SELECT SOIL. MAP

FROM SOIL, PARCELS

WHERE VALUATION > 60000 AND OVERLAY (SOILS, PARCELS)

通过对标准 SQL 的扩展来实现空间数据的查询主要优点是：由于是在标准 SQL 基础上进行扩展的，因而保留了 SQL 的风格，便于熟悉 SQL 的用户的掌握，通用性较好，易于与关系数据库连接。

但 Egenhofer(1992)在分析了扩展 SQL 作为空间数据查询语言的特点和局限后认为，对 SQL 扩展并不是空间数据查询的适当方案，其主要原因是：

- SQL 结构很难描述复杂的空间关系查询；
- 简单的表格形式不能作为空间数据的表现形式。

对于空间数据查询语言，最关键的是对空间概念的描述。理想的情况是空间数据查询语言能完全表示人所理解的空间概念，但目前的空间数据查询语言所能理解和表达的空间概念还很有限。在这方面还需要作进一步的研究。

3. 可视化空间查询

可视化查询是指将查询语言的元素，特别是空间关系，用直观

的图形或符号表示。因为对于某些空间概念用二维图形表示比用一维文字语言描述更清晰和更易理解。

可视化查询主要使用图形、图像、图标、符号来表达概念，具有简单、直观、易于使用的特点。例如某些空间操作可用图 8—1 所示的图符表示(Y. C. Lee&F. L. Chin,1995)。

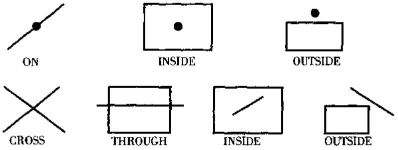


图 8—1 可视化查询

可视化空间查询的主要优点是：自然，直观，易操作，用不同的图符可以组成比较复杂的查询。但也存在一些缺点，如：当空间约束条件复杂时，很难用图符描述；用二维图符表示图形之间的关系时，可能会出现歧义；难以表示“非”关系；不易进行范围(圆、矩形、多边形等)约束；无法进行屏幕定位查询等。

可视化空间查询是为方便用户输入查询条件而设计的，在 GIS 中仍然要翻译成形式化的 SQL 语言。目前可视化空间查询所设计的图符以及所表示的操作缺乏规范性，不能表达所有的空间查询，但其简单、直观的特点值得其他空间查询方法借鉴。

4. 超文本查询

超文本查询把图形、图像、字符等皆当作文本，并设置一些“热点”(HotSpot)，“热点”可以是文本、键等。用鼠标点击“热点”后，可以弹出说明信息、播放声音、完成某项工作等。但超文本查询只能预先设置好，用户不能实时构建自己要求的各种查询。

5. 自然语言空间查询

在空间数据查询中引入自然语言可以使查询更轻松自如。在GIS中很多地理方面的概念是模糊的，例如地理区域的划分实际上并没有象境界一样明确的界线。而空间数据查询语言中使用的概念往往都是精确的。

为了在空间查询中使用自然语言，必须将自然语言中的模糊概念量化为确定的数据值或数据范围。例如查询高气温的城市时，引入自然语言时可表示为：

```
SELECT    name
FROM      Cities
WHERE     temperature is high
```

如果通过统计分析和计算，以及用模糊数学的方法处理，认为当城市气温大于或等于33.75℃时是高气温。则对上述用自然语言描述的查询操作转换为：

```
SELECT    name
FROM      Cities
WHERE     temperature>=33.75
```

在对自然语言中的模糊概念量化时，必须考虑当时的语义环境。例如，对于不同的地区，城市为"高气温"时的温度是不同的；气温的"高(high)"和人身材的"高(high)"也是不同的；等等。因此，引入自然语言的空间数据查询只能适用于某个专业领域的地理信息系统，而不能作为地理信息系统中的通用数据库查询语言。

6. 查询结果的显示

GIS中的空间数据查询功能不能只是简单的数据查询，即不能只给出查询到的数据，应该以最有效的方式将空间数据显示给用户。例如对于查询到的地理现象的属性数据，有时可以以表格的形式显示，有时可以用统计图表的形式显示，可以根据用户的要求来确定。

空间数据的最佳表示方式是地图，因而空间数据查询的结果最好以属性地图的形式表示出来。但目前把查询的结果制作成专

题地图还需要一个比较复杂的过程。为了方便查询结果的显示，Max(1991,1994)在基于扩展SQL的查询语言中增加了图形表示语言，作为对查询结果显示的表示。具有6种显示环境的参数可选定。

(1)显示方式(the display mode)

有5种显示方式用于多次查询结果的运算：刷新、覆盖、清除、相交和强调。

(2)图形表示(the graphical presentation)

用于选定符号、图案、色彩等。

(3)绘图比例尺(the scale of the drawing)

确定地图显示的比例尺(内容和符号不随比例尺变化)。

(4)显示窗口(the window to be shown)

确定屏幕上显示窗口的尺寸。

(5)相关的空间要素(the spatial context)

显示相关的空间数据，使查询结果更容易理解。

(6)查询内容的检查(the examination of the content)

检查多次查询后的结果。

通过选择这些环境参数可以把查询结果以用户选择的不同的形式显示出来，但离把查询结果以丰富多彩的专题地图显示出来的目标还相差很远。

§8.2 空间数据的统计分析

GIS中空间数据的统计分析是指对GIS地理数据库中的属性数据进行统计分析，这里只介绍基本统计量的计算和常用的统计数据的分类分级算法。

1. 属性数据的集中特征数

(1)频数和频率。将变量$x_i(i=1,2,\cdots,n)$按大小顺序排列，并按一定的间距分组。变量在各组出现或发生的次数称为频数，一般

用 f_i 表示。各组频数与总频数之比叫做频率，按如下公式计算：

$$f = fi / \sum_{i=1}^{n} fi \qquad (i=1,2,\cdots,n) \qquad (8-1)$$

根据大数定理，当 n 相当大时，频率可近似地表示事件的概率。

计算出各组的频率后，就可作出频率分布图。若以纵轴表示频率，横轴表示分组，就可作出频率直方图。用以表示事件发生的频率和分布状况。

(2)平均数。平均数反映了数据取值的集中位置，常以 $\overline{X}$ 表示。对于数据 $x_i(i=1,2,\cdots,n)$ 通常有简单算术平均数和加权算术平均数。

简单算术平均数的计算公式为：

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (8-2)$$

加权算术平均数的计算公式为：

$$\overline{X} = \sum_{i=1}^{n} P_i x_i / \sum_{i=1}^{n} P_i \qquad (8-3)$$

其中 P_i 为数据 x_i 的权值。

(3)数学期望。以概率为权值的加权平均数称为数学期望，用于反映数据分布的集中趋势。计算公式为：

$$E_x = \sum_{i=1}^{n} P_i x_i \qquad (8-4)$$

其中 P_i 为事件发生的概率。

(4)中数。对于数据 X，如果有一个数 x，能同时满足以下两式：

$$\begin{cases} P(X \geq x) \geq \dfrac{1}{2} \\ P(X \leq x) \geq \dfrac{1}{2} \end{cases}$$

则称 x 为数据 X 的中数，记为 M_e。

若 X 的总项数为奇数，则中数为：

$$M_e = X_{\frac{1}{2}(n-1)} \tag{8--5}$$

若 X 的总项数为偶数，则中数为：

$$M_e = \frac{1}{2}(X_{\frac{n}{2}} + X_{\frac{n-1}{2}}) \tag{8—6}$$

(5)众数。众数是具有最大可能出现的数值。如果数据 X 是离散的，则称 X 出现的最大可能性的值 x 为众数；如果 X 是连续的，则以 X 分布的概率密度 P(x)取最大值的 x 为 X 的众数。显然，众数可能不是唯一的。

2. 属性数据的离散特征数

在分析 GIS 的属性数据时，不仅要找出数据的集中位置，而且还要查明这些数据的离散程度，即它们相对于中心位置的程度，同时，还要分析它的变化范围。对于两组数据，可能它们的平均数是一样的，但它们分布在平均数附近的疏密程度却可能不同，即它们的离散程度可能不一样。从统计规律的角度讲，离散程度较小的区域，其平均数的代表性较好，反之则较差。很明显，前述的平均数、数学期望、中数和众数是不可能反映数据的离散程度的，因此需要引入刻划离散程度差异的统计特征数，即极差、离差、方差、标准差、变差系数。

(1)极差。极差是一组数据中最大值与最小值之差，即

$$R = \max\{x_1, x_2, \cdots, x_n\} - \min\{x_1, x_2, \cdots, x_n\} \tag{8--7}$$

(2)离差、平均离差与离差平方和。一组数据中的各数据值与平均数之差称为离差，即

$$d = x_i - \overline{x} \tag{8—8}$$

若把离差求平方和，即得离差平方和，记为

$$d^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{8·-9}$$

若将离差取绝对值，然后求和，再取平均数，得平均离差，记为

$$md=\sum_{i=1}^{n}|x_i-\bar{x}|/n \qquad (8—10)$$

平均离差和离差平方和是表示各数值相对于平均数的离散程度的重要统计量。

(3)方差与标准差。方差是均方差的简称，是以离差平方和除以变量个数求得的，记为σ^2，即：

$$\sigma^2=\sum_{i=1}^{n}(x_i-\bar{x})^2/n \qquad (8—11)$$

标准差是方差的平方根，记为：

$$\sigma=\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2/n} \qquad (8—12)$$

(4)变差系数。变差系数用来衡量数据在时间和空间上的相对变化的程度，它是无量纲的量，记为Cv：

$$C_v=\frac{\sigma}{\bar{X}}\times 100\% \qquad (8—13)$$

其中，σ为标准差，$\bar{X}$为平均数。

3. 统计数据的分类分级

为了把GIS地理数据中的统计数据用专题地图的形式表示出来，通常需要对统计数据进行分类和分级。分类和分级的方法很多，下面介绍两种在统计数据处理中经常用到的算法：系统聚类法和最优分割分级法。

(1)系统聚类法

——系统聚类法的基本思想。系统聚类法是分类数据处理中用得最多的一种方法。其基本思想是：首先是n个样本各自成一类，然后规定类与类之间的距离，选择距离最小的两类合并成一个新类，计算新类与其它类的距离，再将距离最小的两类进行合并，这样每次减少一类，直到达到所需的分类数或所有的样本都归为一类为止。

——统计数据的标准化。每个样本单元有多种变量的原始数据，各种变量的量纲和数量大小是很不一致的，变化的幅度也不一

样。假如直接用原始数据进行计算，就会突出绝对值大的变量的作用，而压低绝对值小的变量的作用。为了给每种变量统一量度，在进行模型的统计计算前，往往需要对原始数据进行标准化变换。通常使用标准差标准化的方法。

如果有 n 个样本，每个样本有 m 个数据，则每个变量可记为 $X_{i,j}$，其中 i=1,2,…n；j=1,2,…m。

标准化后的变量为 $X'_{i,j}$，则：

$$X'_{i,j}=\frac{X_{i,j}-\overline{X}_j}{S_j} \qquad (8—14)$$

式中：$\overline{X}_j$ 为第 j 个变量的平均数；S_j 为第 j 个变量的标准差。

在制图数据处理中，常用 S 代表标准差。当用样本标准差对总体标准差进行估算时，则采用无偏估计值，即：

$$S=[\sum_{i=1}^{n}(X_i-\overline{X})^2/(n-1)]^{\frac{1}{2}} \qquad (8-15)$$

经过标准差标准化后，每种变量的平均值为 0，方差为 1。

为了避免计算标准差，并把变量变换到 0 和 1 的范围内，也可采用极差标准化。这时可采用下面公式：

$$X'_{i,j}=\frac{X_{i,j}-\overline{X}_j}{X_{jmax}-X_{jmin}} \qquad (8—16)$$

式中：X_{jmax} 和 X_{jmin} 为第 j 组变量的最大值和最小值。用此公式标准化后的变量范围在 ±1 之间。

——距离系数。对样本进行分类时，个体之间的相似性程度往往用"距离"来度量。它是将每个样本看成是高维空间的一个点，点与点之间用某种法则规定距离，距离近的点归为一类。

设有 n 个样本，每个样本测得 m 项指标，则每个变量记为 X_{ij}，i=1,2,…n；j=1,2,…m。这里只介绍在标准化变量互不相关时采用的欧氏距离。两样本 i 和 j 的距离记为 d_{ij}，则：

$$d_{ij}=\sqrt{\sum_{k=1}^{m}(X_{ik}-X_{jk})^2} \qquad (8—17)$$

——最短距离法系统聚类。类与类之间的距离有许多种定义方法，这里介绍最常用的最短距离法。在最短距离法中，定义两类之间的距离用两类间最近样本的距离来表示。用 d_{ij}表示样本 i 和 j 的距离，用 G_1，G_2，…表示类，用 D_{pq}表示类 G_p 和类 G_q 的距离，则有：

$$D_{pq}=\min d_{ij}$$
$$i\in G_p, j\in G_q$$

用最短距离法聚类的步骤如下：

①计算每两个样本的距离。由于样本 i 和样本 j 的距离 d_{ij}与样本 j 和样本 i 的距离 d_{ji}是相等的，所以只要计算一个。开始时每个样本自成一类，则类之间的距离 D_{pq}就等于样本之间的距离 d_{pq}。

②找出最小的类间距，设为 D_{pq}，则把 G_p 的 G_q 合并为一个新类，记为 G_r。

③计算新类与其它类的距离。新类 G_r 和某一类 G_k 的距离为 D_{rk}，则 D_{rk}可用下面公式计算出：

$$D_{rk}=min\{D_{pk}\cdot D_{qk}\} \qquad (8—18)$$

④重复第②、③步，直到所有元素都成为一类或达到设定的分类数。

表 8—1

系数 方法	α_p	α_q	β	δ
最短距离法	1/2	1/2	0	−1/2
最长距离法	1/2	1/2	0	1/2
中间距离法	1/2	1/2	$-1/4\leqslant\beta\leqslant 0$	0
重心法	n_p/n_r	n_q/n_r	$-\alpha_p\cdot\alpha_q$	0
类平均法	n_p/n_r	n_q/n_r	0	0
可变类平均法	$(1-\beta)n_p/n_r$	$(1-\beta)n_q/n_r$	<1	0
可变法	$(1-\beta)/2$	$(1-\beta)/2$	<1	0
离差平方和法	$(n_k+n_p)/(n_k+n_r)$	$(n_k+n_q)/(n_k+n_r)$	$-n_k/(n_k+n_r)$	0

⑤系统聚类法类间距离计算的统一公式。除了最短距离法外，还有其它计算类与类之间的距离的方法，它们除了计算类与类之间距离的公式不同外，其并类的步骤是完全相同的。而这些距离计算法可以用一个统一的公式表示，而只是系数不同。这个统一的公式为：

$$D_{kr}^2=\alpha_p D_{kp}^2+\alpha_q D_{kq}^2+\beta D_{pq}^2+\delta\left|D_{kp}^2-D_{kq}^2\right| \qquad (8-19)$$

式中系数 α_p、α_q、β、δ 对不同的方法取不同的值，上表列出了八种方法的参数值。统一的公式为编写程序提供了很大的方便。

(2)最优分割分级法

——最优分割分级法的基本原理。最优分割分级是在有序样本不被破坏的前提下，使其分割的级内离差平方和为最小而级间离差平方和为极大的一种分级方法。它可以用来对有序样本或可变为有序(排序)的样本进行分级。

n 个数据按大小顺序排列后，有(n－1)个"空隙"，如分成 k 个等级，则需(k－1)个分级界线。因此，n 个数据分成 k 级的可能分法有$\binom{n-1}{k-1}$种。对于每种分级，我们可按误差函数公式来计算分级误差的大小，以择其优。为了使级内离差平方和最小且级间离差平方和最大，实际上只需要满足级内离差平方和最小即可。因此误差函数公式定义为各级内数据的离差平方和之和。

设对于按大小顺序排列的 n 个数据分成 k 级，其分法可表示为如下形式：

$\{i_1=1,i_1+1,\cdots,i_2-1\}\{i_2,i_2+1,\cdots,i_3-1\},\cdots\cdots\{i_k,i_k+1,\cdots,n\}$

式中 $i_1=1<i_2<i_3\cdots<i_k<n$，这里为了简单起见，将 X 省去了，即 X_{i1} 记为 i_1。
则误差函数为：

$$E(n,k)=\sum_{j=1}^{k}D(i_j,i_{j+1}-1) \qquad (8-20)$$

式中 $D(i_j,i_{j+1}-1)$ 为第 j 级数据(即从 i_j 到 $i_{j+1}-1$)的离差平

方和。

当 K 较大时，直接求最优分割的计算量很大。而通常使用推导出的误差函数递推公式来求最优分割。递推公式为：

$$E(n,k)=min\{E(j-1,k-1)+D(j,n),k\leqslant j\leqslant n\}$$

当 K＝2 时，上式为：

$$E(n,k)=min\{D(1,j-1)+D(j,n),2\leqslant j\leqslant n\}$$

— 最优分割的计算。首先必须把数据按从小到大的顺序进行排序，然后逐步进行分割（分级）。

①最优二分割的计算。根据递推公式，对 m 个数据进行最优二分割（即分为二级）的误差（记为 $S_m(2,j)$，j 为分级点）为：

$$S_m(2,j)=D(1,j)+D(j+1,m),j=1,2,\cdots,m-1$$

挑出其最小误差，则 j 即为分割点，记为 $\alpha_1(m)$。当 m＝n 时，可得 n 个样本的最优二分割为

$\{X_1,X_2,\cdots,X_j\},\{X_{j+1},\cdots,X_n\}$。

为了进行最优二分割，需计算 m＝n，n－1，…，2 的最优二分割，记为：$S_j(2,\alpha_1(j))$，其中，j＝n，n－1，…，2。$\alpha_1(j)$为相应的分割点。

②最优三分割。根据递推公式，在已知 $S_j(2,\alpha_1(j))$时，m 个数据的最优三分割误差（记为 $S_m(3,\alpha_1(j),j)$。j 为最优三分割的第二个分割点，$\alpha_1(j)$是第一个分割点为：

$$S_m(3,\alpha_1(i,j),j)=S_j(2,\alpha_1(j))+D(j+1,n),j=2,3,\cdots,m-1$$

式中 $S_j(2,\alpha_1(j))$在最优二分割时已全部计算出，现只要计算 D(j＋1，n)即可得不同分割点时最优三分割点的第二个分割点，相应的 $\alpha_1(j)$则为第一个分割点。

当 m＝n 时，就是对 n 个样本数据的最优三分割（分三级）。

为了进一步的分级，需计算 m＝n，n－1，…，3 时数据的最优三分割，记为 $S_j[3,\alpha_1(j),\alpha_2(j)]$，其中 j＝n，n－1，…，3。$\alpha_1(j)$和 $\alpha_2(j)$为相应的分割点。

③最优 K 分割。在已完成（K－1）的分割的基础上，用相似的

方法可求出最优 K 分割。

按最优分割分级法计算出的分级结果在用于地图制图前，必须把分级界线的零碎值转换成凑整值。凑整时分级界线的起点和终点分别向较小和较大的数凑整。

§8.3 数字高程模型分析

1. 基于 DEM 的信息提取

(1)坡度的计算。地表单元的坡度就是其法向量$\vec{n}$与 Z 轴的夹角，如图 8—2 所示。坡度 G 的计算公式为：

$$tgG=\sqrt{(\Delta Z/\Delta x)^2+(\Delta Z/\Delta y)^2} \qquad (8—21)$$

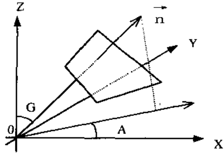


图 8—2 坡度示意图

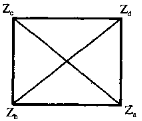


图 8—3 格网 DEM 求坡度

例如，对于格网 DEM，如图 8—3，若 Z_a、Z_b、Z_c、Z_d 是一个格网上的四个格网点的高程，ds 为格网的边长，则格网的坡度可由下式计算：

$$G=\mathrm{arctg}\sqrt{u^2+v^2} \qquad (8—22)$$

$$u=\frac{\sqrt{2}(Z_a-Z_b)}{2ds}$$

$$v=\frac{\sqrt{2}(Z_c-Z_d)}{2ds}$$

若需求格网点上的坡度时，可取 3×3 的格网单元进行类似的

计算。也可求出该格网点八个方向上的坡度，再取其最大值。

在计算出各地表单元的坡度后，可对不同的坡度设定不同的灰度级，或绘出等值线，即可得到坡度图。

(2)坡向的计算。坡向是地表单元的法向量在OXY平面上的投影与X轴之间的夹角，如图8—2。坡向通常要换算成正北方向起算的角度。其计算公式为：

$$\mathrm{tg}A=\frac{\Delta/Z\Delta y}{\Delta Z/\Delta x} \qquad (-\pi<A<\pi) \qquad (8-23)$$

对于格网DEM，如图8—3，则坡度的计算公式为：

$$A=\mathrm{arctg}\left(-\frac{v}{u}\right) \qquad (8-24)$$

$$u=\frac{\sqrt{2}(Z_a-Z_b)}{2ds}$$

$$v=\frac{\sqrt{2}(Z_c-Z_d)}{2ds}$$

在计算出每个地表单元的坡向后，可制作坡向图，通常把坡向分为东、南、西、北、东北、西北、东南、西南8类，再加上平地，共9类，并用不同的色彩显示，即可得到坡向图。

2. 基于DEM的可视化分析

(1)剖面分析。研究地形剖面，常常以线代面，研究区域的地貌形态、轮廓形状、地势变化、地质构造、斜坡特征、地表切割强度等等。如果在地形剖面上叠加上其他地理变量，例如坡度、土壤、植被、土地利用现状等，可以提供土地利用规划、工程选线和选址等的决策依据。

剖面图的绘制应在格网DEM或三角网DEM上进行。已知两点的坐标A(x_1,y_1)，B(x_2,y_2)，则可求出两点连线与格网或三角网的交点，以及各交点之间的距离。然后按选定的垂直比例尺和水平比例尺，按距离和高程绘出剖面图，如图8—4所示：

在绘制剖面图时，需进行高程的插值。在求起始点和终止点A和B的高程时，对于格网DEM可通过其周围的4个格网点内插

出，对于三角网 DEM 可通过该点所在的三角形的三个顶点进行内插。内插的方法可任选。例如可选择距离加权法。则内插点的高程为：

$$Z=\frac{\sum_{i=1}^{n}(Z_i/d_i^2)}{\sum_{i=1}^{n}(1/d_i^2)} \qquad (8—25)$$

其中，对格网 DEM，取 n＝4，对三角网 DEM，取 n＝3，Z_i 为数据点的高程，d_i 为数据点到内插点的距离。

计算格网或三角网交点的高程通常可采用简单的线性内插算法。如图 8—5，格网两点或三角形一条边上的两点为 $A(x_1,y_1,z_1)$、$B(x_2,y_2,z_2)$，交点 C 的坐标为 $C(x_0,y_0,z_0)$，则可计算出 AC 的距离 S_1，AB 的距离 S_2，则 C 点的高程 z_0 为：

$$z_0=\frac{z_2-z_1}{S_2}\cdot S_1 \qquad (8—26)$$

剖面图不一定必须沿直线绘制，也可沿一条曲线绘制。但其绘制方法仍然是相同的。

(2)通视分析。通视分析是指以某一点为观察点，研究某一区域通视情况的地形分析。通视分析的核心是通视图的绘制。

绘制通视图的基本思路是：如图 8—6，以 O 为观察点，对格网 DEM 或三角网 DEM 上的每个点判断通视与否，通视赋值为 1，不通视赋值为 0。由此可形成属性值为 0 和 1 的格网或三角网。对此以 0.5 为值追踪等值线，即得到以 O 为观察点的通视图。因此，判断格网或三角网上的某一点是否通视成为关键。以格网 DEM 为例，如图 8—6，$O(x_o,y_o,z_o)$ 为观察点，$P(x_p,y_p,z_p)$ 为某一格网点，OP 与格网的交点为 A、B、C，则可绘出 OP 的剖面图，如图 8—7。

OP 的倾角 α 可由下式计算出：

$$\mathrm{tg}\alpha=\frac{z_p-z_0}{\sqrt{(x_p-x_0)^2+(y_p-y_0)^2}}$$

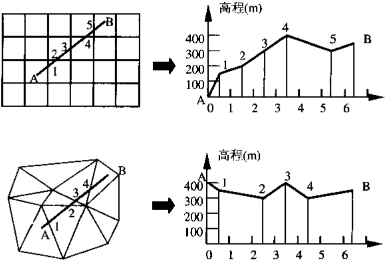


图 8—4 剖面图绘制的示意图

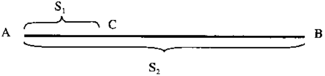
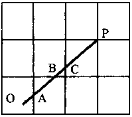


图 8—5 线性内插交点的高程



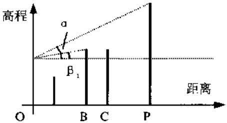


图 8—6 绘制通视图基本思路　　　　图 8—7 利用剖面图判断通视

观察点与各交点的倾角 $\beta_i(i=A,B,C)$ 可由下式计算出：

$$tg\beta_i=\frac{z_i-z_o}{\sqrt{(x_i-x_o)^2+(y_i-y_o)^2}} \qquad i=A、B、C \qquad (8\text{-}27)$$

若 $tg\alpha > \max(tg\beta_i, i = A、B、C)$，则 OP 通视，否则不通视。

三角网 DEM(TIN)中各离散点的通视判断与上述方法类似，也需要通过剖面图来判断。

另一种利用 DEM 绘制通视图的方法是，以观察点 O 为轴，以一定的方位角间隔算出 0°～360°的所有方位线上的通视情况。对于每条方位线，通视的地方绘线，不通视的地方断开，或相反。这样可得出射线状的通视图。其判断通视与否的方法与前述类似。

§8.4 空间数据的叠置分析

1. 基于矢量数据的叠置分析

叠置分析是将同一地区的两组或两组以上的要素进行叠置，产生新特征的分析方法，如图 8—8。叠置的直观概念就是将两幅或多幅地图重叠在一起，产生新多边形和新多边形范围内的属性。

(1)矢量数据叠置的内容。

——点与多边形的叠置。点与多边形的叠置是确定一幅图(或数据层)上的点落在另一幅图(或数据层)的哪个多边形中，这样就可给相应的点增加新的属性内容。

例如，一幅图表示水井的位置，另一幅图表示城市功能分区。两幅图叠置后可得出每个城市功能区(如居住区)有多少水井，也可知道每口水井是位于城市的什么功能区。

点与多边形叠置的算法就是判断点是否在多边形内，可用垂线法或转角法实现。

——线与多边形的叠置。线与多边形的叠置是把一幅图(或一个数据层)中的多边形的特征加到另一幅图(或另一个数据层)的线上。

例如，道路图与境界图叠置，可得到每个政区中各种等级道路的里程。

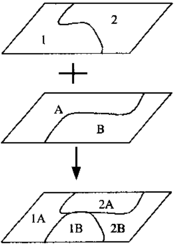


图 8—8 叠置分析示意图

线与多边形叠置的算法就是线的多边形裁剪。算法的具体实现可参考有关计算机图形学的书籍。

——多边形与多边形的叠置。多边形与多边形的叠置是指不同图幅或不同图层多边形要素之间的叠置，通常分为合成叠置和统计叠置。

合成叠置是指通过叠置形成新的多边形，使新多边形具有多重属性，即需进行不同多边形的属性合并。属性合并的方法可以是简单的加、减、乘、除，也可以取平均值、最大值、最小值，或取逻辑运算的结果等。

统计叠置是指确定一个多边形中含有其他多边形的属性类型的面积等，即把其它图上的多边形的属性信息提取到本多边形中来。

例如，土壤类型图与城市功能分区图叠置，可得出商业区中具有不稳定土壤结构的地区有哪些。

多边形与多边形叠置算法的核心是多边形对多边形的裁剪。多边形裁剪比较复杂，因为多边形裁剪后仍然是多边形，而且可能

是多个多边形。多边形裁剪的基本思想是一条边一条边地裁剪。

如图 8—9,多边形 A{a_1,a_2,a_3,a_4,a_5}对多边形 B{$b_1,b_2,b_3,b_4,b_5,b_6,b_7,b_8,b_9,b_{10},b_{11},b_{12}$}裁剪。则可先用 a_1a_2 及其延长线对多边形 B 裁剪,在 a_1a_2 及其延长线上得到交点 P_1、P_2、P_3、P_4、P_5、P_6,则多边形 B 被多边形 A 的 a_1a_2 裁剪后为{$b_1,b_2,b_3,b_4,b_5,P_3,P_2,b_7,b_8,P_1,P_4,b_{11},P_5,P_6$}。当用多边形 A 的每一边顺序对 B 进行裁剪后,就得到了 B 被 A 裁剪后的结果。

但这样构成的被裁剪后的多边形,实际上是一个多边形。因而当一个多边形被裁剪分为几个多边形时,用这种方法就会形成一个在边界有重叠的多边形。虽然对多边形的填充算法没有影响,但有时仍然要把它分成各自独立的多边形。为此,要对被裁剪的多边形定义方向,在计算出每个多边形之间的交点时,都要判断它是出点还是入点,即是从裁剪多边形中出去还是由外边进来,并建立起出入点的索引表。然后根据被裁剪后多边形的点的顺序,从一个出点出发,找与之最近的入点,再从这个入点开始,找出从此点开始的内部线,到一个出点结束,并判断多边形是否闭合。若闭合,则已形成一个多边形,继续进行即可;若不闭合,再找与之最近的入点。直到所有的内部线都使用过一次。多边形裁剪的具体细节请参阅计算机图形学的有关部分。

(2)多边形叠置的位置误差。进行多边形叠置的往往是不同类型的地图,甚至是不同比例尺的地图,因此,同一条边界的数据往往不同,这时在叠置时就会产生一系列无意义的多边形,如图 8—10。而且边界位置越精确,越容易产生无意义多边形。手工方法叠置时可用制图综合来处理无意义的多边形,而计算机处理时则比较复杂,常用如下三种方法:

——在屏幕上显示多边形叠加的情况,人机交互地把小多边形合并到大多边形中。

——确定无意义多边形的面积临界值,把小于临界值的多边形合并到相邻的大多边形中。

——先拟合出一条新的边界线，然后进行叠置操作。

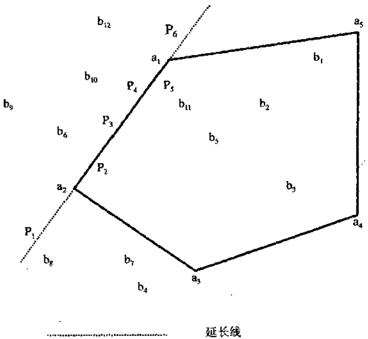


图 8--9　多边形裁剪示意图

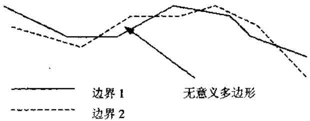


图 8--10　叠置产生的无意义多边形

无论采用哪种方法来处理无意义多边形，都会产生误差。

2. 基于栅格数据的叠置分析

(1)单层栅格数据的分析。

——布尔逻辑运算。栅格数据可以按其属性数据的布尔逻辑运算来检索，即这是一个逻辑选择的过程。布尔逻辑为AND、OR、XOR、NOT，如图8--11所示。

例如可以用条件(A AND B) OR C进行检索。其中，A为土壤是粘性的，B为PH值>7.0的，C为排水不良的。这样就可把栅格数据中土壤结构为粘性的、土壤PH值大于7.0的、或者排水不良的区域检索出来。

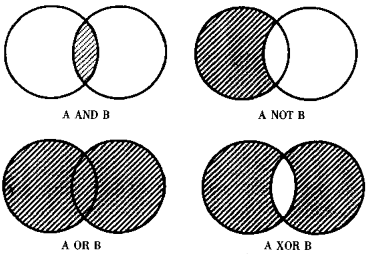


图8—11　布尔逻辑运算

布尔逻辑运算可以组合更多的属性作为检索条件，例如加上面积和形状等条件，以进行更复杂的逻辑选择运算。

——重分类。重分类是将属性数据的类别合并或转换成新类。即对原来数据中的多种属性类型，按照一定的原则进行重新分类，以利于分析。在多数情况下，重分类都是将复杂的类型合并成简单的类型。例如，可以将各种土壤类型重分类为水面和陆地两种类型。

重分类时必须保证多个相邻接的同一类别的图形单元应获得

一个相同的名称，并且这些图形单元之间的边应该去掉，从而形成新的图形单元。如图 8—12 所示。

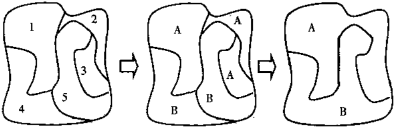


图 8—12 重分类的过程

——滤波运算。对栅格数据的滤波运算是指通过一移动的窗口（如 3×3 的像元），对整个栅格数据进行过滤处理，使窗口最中央的像元的新值定义为窗口中像元值的加权平均值。

栅格数据的滤波运算可以将破碎的地物合并和光滑化，以显示总的状态和趋势，也可以通过边缘增强和提取，获取区域的边界。

——特征参数计算。对栅格数据可计算区域的周长、面积、重心等，以及线的长度、点的坐标等。

在栅数数据上量算面积有其独特的方便之处，只要对栅格进行计数，再乘以栅格的单位面积即可。

在栅格数据中计算距离时，距离有四方向距离、八方向距离、欧几里德距离等多种意义。四方向距离是通过水平或垂直的相邻像元来定义路径的；八方向距离是根据每个像元的八个相邻像元来定义的；在计算欧几里德距离时，需将连续的栅格线离散化，再用欧几里德距离公式计算。

对图 8—13 中的线，用四方向距离计算的距离为 6，用八方向计算的距离为 $2+2*\sqrt{2}$。

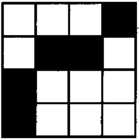


图 8-13　栅格数据的距离计算

——相似运算。相似运算是指按某种相似性度量来搜索与给定物体相似的其他物体的运算。

(2)多层栅格数据的叠置分析。叠置分析是指将不同图幅或不同数据层的栅格数据叠置在一起，在叠置地图的相应位置上产生新的属性的分析方法。新属性值的计算可由下式表示：

$U=f(A,B,C,\cdots\cdots)$

其中，A，B，C 等表示第一、二、三等各层上确定的属性值，f 函数取决于叠置的要求。

多幅图叠置后的新属性可由原属性值的简单的加、减、乘、除、乘方等计算出，也可以取原属性值的平均值、最大值、最小值或原属性值之间逻辑运算的结果等，甚至可以由更复杂的方法计算出，如新属性的值不仅与对应的原属性值相关，而且与原属性值所在的区域的长度、面积、形状等特性相关。

栅格叠置的作用可归纳为：

——类型叠置。即通过叠置获取新的类型。如土壤图与植被图叠置，以分析土壤与植被的关系。

——数量统计。即计算某一区域内的类型和面积。如行政区划图和土壤类型图叠置，可计算出某一行政区划中的土壤类型数，以及各种类型土壤的面积。

——动态分析。即通过对同一地区的相同属性、不同时间的栅格数据的叠置，分析由时间引起的变化。

——益本分析。即通过对属性和空间的分析，计算成本、价值等。

——几何提取。即通过与所需提取的范围的叠置运算，快速地进行范围内信息的提取。

在进行栅格叠置的具体运算时，可以直接在未压缩的栅格矩阵上进行，也可在压缩编码(如游程编码、四叉树编码)后的栅格数据上进行。它们之间的差别主要在于算法的复杂性、算法的速度、所占用的计算机内存等。

§8.5 空间数据的缓冲区分析

1. 基于矢量数据的缓冲区分析

(1)缓冲区及其作用。在这里，缓冲区的概念与计算机技术中的缓冲区概念无关，而是指在点、线、面实体的周围，自动建立的一定宽度的多边形，如图 8—14。

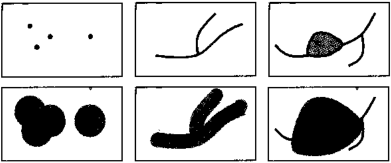


图 8—14 点线面的缓冲区

缓冲区分析是 GIS 的基本空间操作功能之一。例如，某地区有危险品仓库，要分析一旦仓库爆炸所涉及的范围，这就需要进行点缓冲区分析；如果要分析因道路拓宽而需拆除的建筑物和需搬

迁的居民，则需进行线缓冲区分析；而在对野生动物栖息地的评价中，动物的活动区域往往是在距它们生存所需的水源或栖息地一定距离的范围内，为此可用面缓冲区进行分析，等等。

在建立缓冲区时，缓冲区的宽度并不一定是相同的，可以根据要素的不同属性特征，规定不同的缓冲区宽度，以形成可变宽度的缓冲区。例如，沿河流绘出的环境敏感区的宽度应根据河流的类型而定。这样就可根据河流属性表，确定不同类型的河流所对应的缓冲区宽度，以产生所需的缓冲区。

(2)缓冲区的建立。点的缓冲区建立时，只需要给定半径绘圆即可。面的缓冲区只朝一个方向，而线的缓冲区需在线的左右配置。下面简介线的缓冲区的建立思路。

在建立线缓冲区时，通常首先要对线进行化简，以加快缓冲区建立的速度。这种对线的化简称为线的重采样。具体的算法设计可采用线的矢量数据压缩算法，如§7.4中介绍的方法。

建立线缓冲区就是生成缓冲区多边形。只需在线的两边按一定的距离(缓冲距)绘平行线，并在线的端点处绘半圆，就可连成缓冲区多边形。

对一条线所建的缓冲区有可能重叠，如图8—15。这时需把重叠的部分去除。基本思路是，对缓冲区边界求交，并判断每个交点是出点还是入点，以决定交点之间的线段保留或删除。这样就可得到岛状的缓冲区。

在对多条线建立缓冲区时，可能会出现缓冲区之间的重叠。这时需把缓冲区内部的线段删除，以合并成连通的缓冲区(图8—16)。

2. 基于栅格数据的缓冲区分析

缓冲区分析在GIS中用得较多，但对矢量数据的缓冲区操作比较复杂，而在栅格数据中可看作是对空间实体向外进行一定距离的扩展，因而算法比较简单。



(1)输入数据　　(2)缓冲区操作　　(3)重叠处理后的缓冲区

图 8—15　线缓冲区重叠的处理



(1)输入数据　　(2)缓冲区操作　　(3)重叠处理后的缓冲区

图 8—16　多条线缓冲区重叠的处理

§8.6　泰森多边形分析

1. 泰森多边形及其特性

荷兰气象学家泰森（A・H・Thiessen）提出了一种根据离散分布的气象站的降雨量来计算平均降雨量的方法，即将所有相邻气象站连成三角形，作这些三角形各边的垂直平分线，于是每个气象站周围的若干垂直平分线便围成一个多边形。用这个多边形内所包含的一个唯一气象站的降雨强度来表示这个多边形区域内的降雨强度，并称这个多边形为泰森多边形。如图 8—17，其中虚线构成的多边形就是泰森多边形。泰森多边形每个顶点是每个三角形的外接圆圆心。泰森多边形也称为 Voronoi 图，或 dirichlet 图。

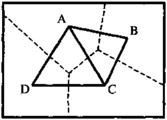


图 8-17　构建泰森多边形

泰森多边形的特性是：

・每个泰森多边形内仅含有一个离散点数据；

・泰森多边形内的点到相应离散点的距离最近；

・位于泰森多边形边上的点到其两边的离散点的距离相等。

泰森多边形可用于定性分析、统计分析、邻近分析等。例如，可以用离散点的性质来描述泰森多边形区域的性质；可用离散点的数据来计算泰森多边形区域的数据；判断一个离散点与其他哪些离散点相邻时，可根据泰森多边形直接得出，且若泰森多边形是 n 边形，则就与 n 个离散点相邻；当某一数据点落入某一泰森多边形中时，它与相应的离散点最邻近，无需计算距离。

在泰森多边形的构建中，首先要将离散点构成三角网。这种三角网称为 Delaunay 三角网。

2. Delaunay 三角网的构建

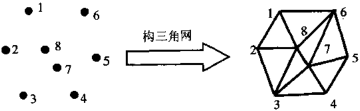


图 8—18　由离散数据点构建三角网

Delaunay 三角网的构建也称为不规则三角网的构建，就是由离散数据点构建三角网，如图 8—18，即确定哪三个数据点构成一

个三角形，也称为自动联接三角网。即对于平面上 n 个离散点，其平面坐标为(x_i,y_i)，$i=1,2,\cdots,n$，将其中相近的三点构成最佳三角形，使每个离散点都成为三角形的顶点。自动联接三角网的结果为所有三角形的三个顶点的标号，如：

1,2,8

2,8,3

3,8,7

⋮

为了获得最佳三角形，在构三角网时，应尽可能使三角形的三内角均成锐角，即符合 Delaunay 三角形产生的准则：

——任何一个 Delaunay 三角形的外接圆内不能包含任何其他离散点。

——相邻两个 Delaunay 三角形构成凸四边形，在交换凸四边形的对角线之后，六个内角的最小者不再增大。该性质即为最小角最大准则。

下面介绍 Tsai (1993) 提出的在 n 维欧拉空间中构造 Delaunay 三角形的通用算法——凸包插值算法。

(1)凸包生成。

①求出点集中满足 min(x−y)、min(x+y)、max(x−y)、max(x+y)的四个点，并按逆时针方向组成一个点的链表。这 4 个点是离散点中与包含离散点的外接矩形的 4 个角点最近的点。这 4 个点构成的多边形作为初始凸包。

②对于每个凸包上的点 I，设它的后续点为 J，计算矢量线段 IJ 右侧的所有点到 IJ 的距离，求出距离最大的点 K。

③将 K 插入 I，J 之间，并将 K 赋给 J。

④重复②、③)步，直到点集中没有在线段 IJ 右侧的点为止。

⑤将 J 赋给 I，J 取其后续点，重复②、③、④步。

⑥当凸包中任意相邻两点连线的右侧不存在离散点时，结束点集凸包求取过程。

完成这一步后，形成了包含所有离散点的多边形(凸包)，如图8—19所示。

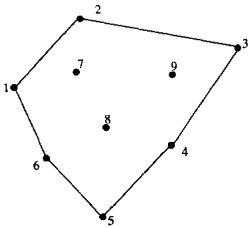


图8-19 凸包生成

(2)环切边界法凸包三角剖分。在凸包链表中每次寻找一个由相邻两条凸包边组成的三角形，在该三角形的内部和边界上都不包含凸包上的任何其它点。将这个点去掉后得到新的凸包链表。重复这个过程，直到凸包链表中只剩三个离散点为止。将凸包链表中的最后三个离散点构成一个三角形，结束凸包三角剖分过程。

完成这一步后，将凸包中的点构成了若干 Delaunay 三角形，如图8—20所示。

(3)离散点内插。在对凸包进行三角剖分之后，不在凸包上的其余离散点，可采用逐点内插的方法进行剖分。基本过程为：

①找出外接圆包含待插入点的所有三角形，构成插入区域。

②删除插入区域内的三角形公共边，形成由影响三角形顶点构成的多边形。

③将插入点与多边形所有顶点相连，构成新的 Delaunay 三角形。

④重复①、②、③，直到所有非凸壳离散点都插入完为止。

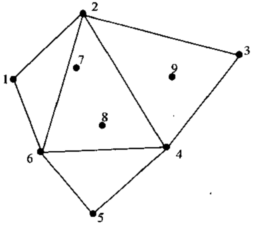


图 8 20 凸包三角部分

完成这--步后，就完成了 Delaunay 三角网的构建，如图 8—21 所示。

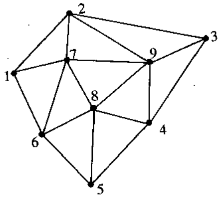


图 8—21 离散点内插

3. 泰森多边形的建立

建立泰森多边形算法的关键是对离散数据点合理地连成三角网，即构建 Delaunay 三角网。建立泰森多边形的步骤为：

①离散点自动构建三角网，即构建 Delaunay 三角网。对离散

点和形成的三角形编号，记录每个三角形是由哪三个离散点构成的。

②找出与每个离散点相邻的所有三角形的编号，并记录下来。这只要在已构建的三角网中找出具有一个相同顶点的所有三角形即可。

③对于每个离散点相邻的三角形按顺时针或逆时针方向排序，以便下一步连接生成泰森多边形。排序的方法如图 8—22 所示。设离散点为 o。找出以 o 为顶点的一个三角形，设为 A；取三角形 A 除 o 以外的另一顶点，设为 a，则另一个顶点也可找出，即为 f；则下一个三角形必然是以 of 为边的，即为三角形 F；三角形 F 的另一顶点为 e，则下一三角形是以 oe 为边的；如此重复进行，直到回到 oa 边。

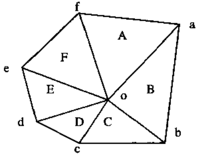


图 8--22　三角形排序

④计算每个三角形的外接圆圆心，并记录之。

⑤根据每个离散点的相邻三角形，连接这些相邻三角形的外接圆圆心，即得到泰森多边形。对于三角网边缘的泰森多边形，可作垂直平分线与图廓相交，与图廓一起构成泰森多边形。

§8.7　空间数据的网络分析

1. 网络图论基础

网络分析是GIS空间分析的重要组成部分。在网络分析中用到的网络模型是数学模型中离散模型的一部分。分析和解决网络模型的有力工具是图论。

图论中的“图”并不是通常意义下的几何图形或物体的形状图,而是一个以抽象的形式来表达确定的事物,以及事物之间具备或不具备某种特定关系的数学系统。

由点集合 V 和点与点之间的连线的集合 E 所组成的集合对 (V,E)称为图,用 G(V,E)来表示。V 中的元素称为节点,E 中的元素称为边。节点集 V 与边集合 E 均为有限的图称为有限图。本章只讨论有限图。

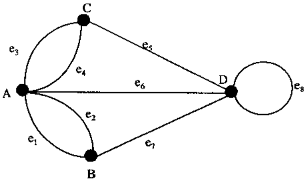


图 8--23 “图”的节点集合和边集合

在图 8—23 中,节点集合 V={A,B,C,D},边集合为 E={e1,e2,e3,e4,e5,e6,e7,e8}。连接两个节点间的边可能不止一条,如 e1,e2 都连接 A 和 B。连接同一节点的边称为自圈,如 e8。如不特别声明,本章不讨论具有自圈和多重边的图。

如果图中的边是有向的,则称为有向图,如图 8—24 所示。在无向图中,首尾相接的一串边的集合称为路。在有向图中,顺向的首尾相接的一串有向边的集合称为有向路。通常用顺次的节点或边来表示路或有向路。如图 8—25 中,{e1,e2,e4}为一条路,该路也可用{v1,v2,v3,v5}来表示。

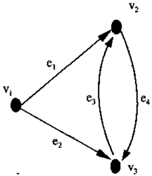


图 8—24

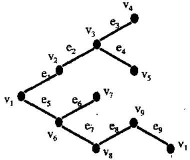


图 8—25 有向图的有向路和树

起点和终点为同一节点的路称为回路(或圈)。如果一个图中，任意两个节点之间都存在一条路，称这种图为连通图。若一个连通图中不存在任何回路，则称为树，如图 8—25。由树的定义，直接得出下列性质：

- 树中任意两节点之间至多只有一条边；
- 树中边数比节点数少 1；
- 树中任意去掉一条边，就变成不连通图；
- 树中任意添一条边，就会构成一个回路；

任意一个连通图，或者是树，或者去掉一些边后形成树，这种树称为这个连通图的生成树。一般来说，一个连通图的生成树可能不止一个。

如果图中任一边(i,j)都赋一个数 ω(i,j)，称这种数为该边的权数。赋以权数的图成为赋权图。有向图的各边赋以权数后，成为有向赋权图。赋权图在实际问题中非常有用。根据不同的实际情况，权数的含义可以各不相同。例如，可用权数代表两地之间的实际距离或行车时间，也可用权数代表某工序所需的加工时间等。

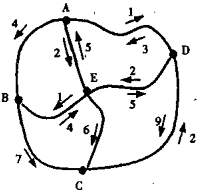
2. 路径分析

(1)最短路径分析。在路径分析中最典型、最常用的是最短路径的选择。在最短路径选择中，两点之间的距离可以定义为实际的

距离，也可定义为两点间的时间、运费、流量等，换句话说，可定义为使用这条边所需付出的代价。因此，可以对不同的专题内容进行最短路径分析。下面介绍的最短路径搜索的算法是狄克斯特拉(Dijkstra)在 1959 年提出的，被公认为是最好的算法之一。

——距离矩阵的计算。GIS 中的网络可以看作是图，可以是有向图，也可以是无向图。对于无向图，可当作有向图来处理。

为了求出最短路径，需先计算两点间的距离，并形成距离矩阵。若两点间没有路，则距离为∞。图 8—26 为网络图及其距离矩阵。



	A	b	C	D	E
A	0	4	∞	1	2
B	∞	0	7	∞	4
C	∞	∞	0	2	∞
D	3	∞	9	0	2
E	5	1	6	5	0

图 8- 26　网络图和距离矩阵

——最短路径搜索的依据。网络图中的最短路径应该是一条简单路径，即是一条不与自身相交的路径。

最短路径搜索的基本依据是，若从点 S 到点 T 有一条最短路径，则该路径上的任何点到 S 的距离都是最短的。证明从略。

为了进行最短路径搜索，令 d(X,Y)表示点 X 到 Y 的距离，D(X)表示 X 到起始点 S 的最短距离。在下列搜索算法中，还需假定两点之间的距离不为负。

——最短路径搜索的步骤：

• 对起始点 S 作标记，且对所有顶点令 D(X)=∞，Y=S。

· 对所有未作标记的点按以下公式计算距离，

$D(X)=min\{D(X),d(Y,X)+D(Y)\}$

其中 Y 是最后一个作标记的点。

取具有最小值的 D(X)，并对 X 作标记，令 Y=X。

若最小值的 D(X)为∞，则说明 S 到所有未标记的点都没有路，算法终止；否则继续。

· 如果 Y 等于 T，则已找到 S 到 T 的最短路径，算法终止；否则转到上一步继续搜索。

例如，对图 8—26，需搜索 A 到 C 的最短路径，则：

· 对 A 作标记，按公式计算所有标记点的距离。

结果为 D(B)=4，D(C)=∞，D(D)=1，D(E)=2。

最小值为 D(D)=1。

· 对 D 作标记，按公式算 D(B)、D(C)、D(E)。

$D(B)=min\{D(B),d(D,B)+D(D)\}$

$=min\{4,\infty+1\}=4$

$D(C)=min\{D(C),d(D,C)+D(D)\}$

$=min\{\infty,9+1\}=10$

$D(E)=min\{D(E),d(D,E)+D(D)\}$

$=min\{2,2+1\}=2$

· 对 E 作标记，计算 D(B)，D(C)。

$D(B)=min\{D(B),d(E,B)+D(E)\}$

$=min\{4,1+2\}=3$

$D(C)=min\{D(C),d(E,C)+D(E)\}$

$=min\{10,6+2\}=8$

最小值为 D(B)=3。

· 对 B 作标记，计算 D(C)

$D(C)=min\{D(C),d(B,C)+D(B)\}$

$=min\{8,7+3\}=8$

· 根据顺序记录的标记点，以及最小值的取值情况，可得到最

短路径为 A→E→C，最短距离为 8。

(2)最小生成树。生成树是图的极小连通子图。一个连通的赋权图 G 可能有很多的生成树。设 T 为图 G 的一个生成树，若把 T 中各边的权数相加，则这个和数称为生成树 T 的权数。在 G 的所有生成树中，权数最小的生成树称为 G 的最小生成树。

在实际应用中，常有类似在 n 个城市间建立通信线路这样的问题。这可用图来表示，图的顶点表示城市，边表示两城市间的线路，边上所赋的权值表示代价。对 n 个顶点的图可以建立许多生成树，每一棵树可以是一个通信网。若要使通信网的造价最低，就需要构造图的最小生成树。

构造最小生成树的依据有两条：

——在网中选择 n－1 条边连接网的 n 个顶点；

——尽可能选取权值为最小的边。

下面介绍构造最小生成树的克罗斯克尔(Kruskal)算法。该算法是 1956 年提出的，俗称"避圈"法。设图 G 是由 m 个节点构成的连通赋权图，则构造最小生成树的步骤如下：

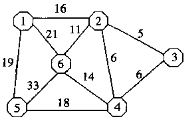
①先把图 G 中的各边按权数从小到大重新排列，并取权数最小的一条边为 T 中的边。

②在剩下的边中，按顺序取下一条边。若该边与 T 中已有的边构成回路，则舍去该边，否则选进 T 中。

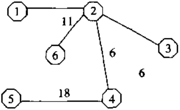
③重复②，直到有 m－1 条边被选进 T 中，这 m－1 条边就是 G 的最小生成树。

设有如图 8—27(1)所示的图，图的每条边上标有权数。为了使权数的总和为最小，应该从权数最小的边选起。在此，选边(2，3)；去掉该边后，在图中取权数最小的边，此时，可选(2，4)或(3，4)，取(2，4)；去掉(2，4)边，下一条权数最小的边为(3，4)，但使用边(3，4)后会出现回路，故不可取，应去掉边(3，4)；下一条权数最小的边为(2，6)；依上述方法重复，可形成图 8—27(2)所示的最小生成树。如果前面不取(2，4)，而取(3，4)，则形成图 8—27(3)所示

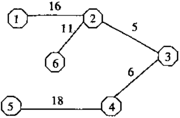
的最小生成树。



(1)赋权图



(2)最小生成树之一



(3)最小生成树之二

图 8—27 图的最小生成树

3. 最小费用最大流

在地理网络中进行着物质和能量的流动,形成各种各样的流。

设有一个水管网络,只有一个进水口和一个出水口。每个管道用其截面积作为权数,用于反映单位时间内可能通过的最大流量(称为容量)。有稳定水流注入进水口,经过网络从出水口流出。这样的一个稳定的流动称为“流”,具有如下性质:

- 流是有向的;
- 管道的流量不可能超过最大流量;
- 每个内部节点处流向和流出节点的流量相等;
- 进水口的流量等于出水口的流量。

网络流的基本问题为:设一个有向赋权图 G(V,E),V={s,a,b,c,…,s'},其中有两个特殊的节点 s 和 s'。s 称为发点,s' 称为收点。图中各边的方向和权数表示允许的流向和最大可能的流量

(容量)。问在这个网络图中从发点流出到收点汇集，最大可通过的实际流量为多少？流向的分布情况为怎样？

最大流问题讨论的是，在一个地理网络中怎样安排网上的流，使从发点到收点的流量最大。在实际应用中，不仅要考虑使网络上的流量最大，而且要使运送流的费用或代价最小。这就是最小费用最大流量问题。

设有一个网络图 G(V,E)，$V=\{s,a,b,c,\cdots,s'\}$，E 中的每条边(i,j)对应一个容量 c(i,j)与输送单位流量所需费用 a(i,j)。如有一个运输方案(可行流)，流量为 f(i,j)，则最小费用最大流问题就是这样一个求极值问题：

$$\min_{f\in F} d(f)=\min_{f\in F}\sum_{(i,j)\in E} a(i,j)f(i,j) \tag{8—28}$$

其中 F 为 G 的最大流的集合，即在最大流中寻找一个费用最小的最大流。

确定最小费用最大流的过程实际上是一个多次迭代的过程。基本思想是：从零流为初始可行流开始，在每次迭代过程中对每条边赋予与 c(i,j)(容量)、a(i,j)(单位流量运输费用)、f(i,j)(现有流的流量)有关的权数 $\omega(i,j)$，形成一个有向赋权图。再用求最短距离路径的方法确定由发点 s 至收点 s' 的费用最小的非饱和路，沿着该路增加流量，得到相应的新流。经过多次迭代，直至达到最大流为止。

构造权数的方法如下：

对任意边(i,j)，根据现有的流 f，该边上的流量可能增加，也可能减少。因此，每条边赋予向前费用权 $\omega^+(i,j)$与向后费用权 ω (i,j)：

$$\omega^+(i,j)=\begin{cases} a(i,j), & 若\ f(i,j)<c(i,j) \\ +\infty, & 若\ f(i,j)=c(i,j) \end{cases}$$

$$\omega^-(i,j)=\begin{cases} -a(i,j), & 若\ f(i,j)>0 \\ +\infty, & 若\ f(i,j)=0 \end{cases}$$

对于赋权后的有向图，如把权 $\omega(i,j)$ 看作长度，即可确定 s 到 s' 的费用最小的非饱和路，等价于从 s 到 s' 的最短路。确定了非饱和路后，就可确定该路的最大可增流量。因此需对每一条边确定一个向前可增流量 $\triangle^+(i,j)$ 与向后可增流量 $\triangle^-(i,j)$：

$$\Delta^+(i,j)=\begin{cases}c(i,j)-f(i,j), & 若\ f(i,j)<c(i,j)\\ 0, & 若\ f(i,j)=c(i,j)\end{cases}$$

$$\Delta^-(i,j)=\begin{cases}f(i,j), & 若\ f(i,j)>0\\ 0, & 若\ f(i,j)=0\end{cases}$$

因此，确定最小费用最大流的具体算法如下：

①从零流开始，令 $f=0$

②赋权

$$当\ f(i,j)<c(i,j),\begin{cases}\omega^+(i,j)=a(i,j)\\ \Delta^+(i,j)=c(i,j)-f(i,j)\end{cases}$$

$$当\ f(i,j)=c(i,j),\begin{cases}\omega^+(i,j)=+\infty\\ \Delta^+(i,j)=0\end{cases}$$

$$当\ f(i,j)<c(i,j),\begin{cases}\omega^-(i,j)=-a(i,j)\\ \Delta^-(i,j)=f(i,j)\end{cases}$$

$$当\ f(i,j)=0\quad\begin{cases}\omega^-(i,j)=+\infty\\ \Delta^-(i,j)=0\end{cases}$$

③确定一条从 s 到 s' 的最短路

$$R(s,s')=\{(s,i_1),(i_1,i_2),\cdots,(i_k,s')\}$$

若 $R(s,s')$ 的长度为 $+\infty$，表明已得到最小费用最大流，则停止；否则转向④。

④确定沿着该路 $R(s,s')$ 的最大可增流量

$$a=min\{\triangle(s,i_1),\triangle(i_1,i_2),\cdots,\triangle(i_k,s')\}$$

其中根据边的取向决定取 $\triangle^+$ 或 $\triangle$ 。

⑤生成新的流

$$f(i,j)=\begin{cases}f(i,j)+\alpha & (i,j)\text{为向前边}\\ f(i,j)-\alpha & (i,j)\text{为向后边}\end{cases}$$

若 f(i,j)已为最小费用最大流，则停止；否则转向②。

4. 网络上的定位与分配模型的启发式算法

定位与分配模型是根据需求点的空间分布，在一些候选点中选择给定数量的供应点以使预定的目标方程达到最佳结果。不同的目标方程就可以求得不同的结果。在运筹学的理论中，定位与分配模型可用线性规划求得全局性的最佳结果。由于其计算量以及内存需求巨大，所以在实际应用中常用一些启发式算法来逼近或求得最佳结果。

在众多的启发式算法中，交换式算法（Interchange）用得最多，其中著名的有 Teitz—Bart 算法。这一算法主要用来解决 P 中心的定位分配问题。这一问题是在 m 个候选点中选择 P 个供应点为 n 个需求点服务，使得为这几个需求点服务的总距离（或时间或费用）为最少。假设 ω_i 记为需求点 i 的需求量。D_{ij}记为从候选点 j 到需求点 i 的距离，则 P 中心问题可叙为：

$$\min\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}\cdot\omega_i\cdot d_{ij}\right)$$

并满足：$\sum_{j=1}^{m}\alpha_{ij}=1, i=1,\cdots,n$

和

$$\sum_{j=1}^{m}\left(\prod_{i=1}^{n}\alpha_{ij}\right)=p, p<m\leqslant n$$

其中，α_{ij}是分配的指数。如果需求点 i 受供应点 j 的服务，则其值为 1，否则为 0。

$$\alpha_{ij}=\begin{cases}1, i \text{ 由 } j \text{ 服务}\\ 0, \text{其它}\end{cases}$$

上述两个约束条件是为了保证每个需求点仅受一个供应点服务，并且只有 p 个供应点。

Teitz－Bart 算法可以叙述如下：

①先选 p 个候选点作为起始供应点集 P_1:C_1,C_2,……C_p。

②将所有的需求点归为它们最邻近的供应点服务以使其距离为最短。计算总的加权距离为 B_1。

③从未选入的候选点集中选一候选点 C_b,C_b 不在 P_1 之中。

④对 P_1 中的每一供应点 C_j,用 C_b 来代替,并计算其总加权距离的变化$\triangle b_j$。

⑤从 P_1 中找出一个供应点 C_k,其被 C_b 取代可以使总加权距离减少最多,即

$$\triangle b_k < 0 \text{ 并且}$$

$$\triangle b_k = \min \triangle b_j (j=1,\cdots,p)$$

这一C_b 取代 C_k,必须使总加权距离减少。不然就不存在取代。

⑥如果在第⑤步中找到 C_k,则 C_b 取代 P_1 中的 C_k,并改名 P_1 为 P_2。重新计算总的加权距离,记为 B_2。如果在第⑤步中没有找到这样的 C_k,转到第⑦步,保留 P_1。

⑦选另一个未试过的候选点,既不在 P_1 中也不在 P_2 中。重复第④到⑥步。

⑧当所有不在 P_1 中的候选点都试过后,其结果记为 P_t,取代 P_1。并重复第②到⑦步的循环。

⑨每完成一个循环②到⑧,如果没有任何取代能减少总的加权距离,则停止。其最后的结果 P_t 则是所求的 P 个中心的供应点。

在这一算法中,主要的计算是在将需求点分配到其最邻近的供应点,并求总的加权距离。在这多层次的循环计算中,不可能临时求解点间的最小距离,这些数据必须预先求得,随时取用,以提高算法的效率。这些最小距离正是用最短路径算法求得的。采用类似动态数据结构——供应点数据串和需求点数据串。数据串的长度常常可用实际应用中服务距离的最大值来限制,可以大大减少计算时间和内存。从供应点的数据串中,可以非常容易找到哪些

需求点是在该点的服务范围，而从需求点数据串中可以找出其相邻的供应点。这在定位与分配的启发式算法中起到了关键的作用。

实际应用中，在选择这样的供应点时，并不只是要使总的加权距离为最小，有时需要使总的服务范围为最大，有时又限定服务的最大距离不能超过一定的值，例如在城市各街区建立图书馆、医院等公共设施，希望各居民住家能到这些设施的路途最短，而在建立消防站、救护车站时，不仅需要路途最短，而且常常规定到最远的住宅，其路途必须在 2 至 3 分钟以内到达，增加了新的约束条件。而在设计有线电视中转站或电话的中心交换站时，不仅要节省电缆或电话线，而且还要使服务的范围最大，这样可以增加用户，提高收入，因此仅仅是 P 中心模型不足以解决更多的实际问题。

在 GIS 系统设计中，常常也不可能对某一问题设计一个专用的算法。因此对常用算法的扩充与推广，意义更为重大。对上述 P 中心模型的目标方程作如下修改，就可以引伸求解其它类型的问题：

$$\Delta b_j = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} \cdot E_{ij} \qquad (8-29)$$

其中 $E_{ij} = \omega_i \cdot d_{ij}$

在 P 中心问题中，如果希望所有的需求点在一给定的理想的服务距离 S 内，则仅需对 E_{ij} 作出如下修改：

$$E_{ij} = \begin{cases} \omega_i \cdot d_{ij}, d_{ij} \leqslant S \\ P, d_{ij} > S \end{cases}$$

其中，P 为阈值，是一个很大的值。

而对最大服务范围问题（MaximalCovering），并希望需求点是在一给定的服务范围 S 内，则对 E_{ij} 作出如下修改：

$$E_{ij} = \begin{cases} 0, D_{ij} \leqslant S \\ \omega_i, d_{ij} > S \end{cases}$$

从目标方程可以看出，对给定的服务距离内 S 的需求点的服务，并不会增加总的加权距离，这样这一算法尽可能选择供应点以

增加服务量，而对距离 S 之外的需求点的服务也不受距离的直接影响。

如果需要限定服务范围在一给定的最远距离 T，则可以设置如下：

$$E_{ij}=\begin{cases}0, & d_{ij}\leqslant S\\ \omega_i, & S<d_{ij}\leqslant T\\ P, & d_{ij}>T\end{cases}$$

其中 P 为阈值，是一个很大的值。这样会逼使需求点尽可能地被服务于 T 距离以内。

通过类似的更改，我们可以用相同的算法有效地解决不同的问题。

现在问题又回到原来的算法上，Teitz－Bart 算法其计算量仍然很大，因为在其多层次循环取代的计算有很大的计算量，尽管有相应的数据结构支持，对大的网络的计算仍需要相当长的时间。那么，我们是否可以利用空间域的特性，来进一步改进或引入新的算法呢？考虑计算量繁重的第 4 和第 5 步，在用来选入点试验替代已选入 A 点的过程中，是否应该考虑所有点。事实上，利用空间邻近的相关性，这样的取代计算并不需要对所有点进行，而其受影响的点应是其附近的点，这样可以大大减少计算时间。P. Densham 和 G. Rushton 正是利用空间上的这一特性，在其对 Teitz－Bart 算法的修改中，引入了一个新的全局和区域性算法（Global and Regional）。Densham－Rushton 的这一新的算法充分利用候选点和需求点数据的信息，在计算时大大减少了计算量。

Densham－Rushton 算法可以叙述如下：

①先选 P 个候选点作为起始供应点集，并将所有需求点分配到最近的供应点，计算其目标方程值，即总的加权距离。

②作全局性调查

(a)检验所有选择的供应点，选定一个供应点准备删去，它的删去仅引起最小的目标方程值的增加。

(b)从未选入的候选点中，寻找一个候选点来代替(a)中选择的供应点，可以最大限度地减少目标方程值。

(c)如果(b)中选择的点所减少的目标方程值大于(a)中选择的点所增加的目标方程的值，用(b)中点代替(a)中点，并更新目标方程值，并到(a)步重复检验。否则的话，转入③步。

③对每一供应点依次作出区域性调整：

(a)如果不是固定的供应点，用它的邻近的候选点来代替检验。

(b)如果这一代替可以最大程度地减少目标方程值，则进行这一替换，直到 P—1 个供应点都被检验，并无新的替换为止。

④重复第②步和第③步直到两步都无新的替换为止。

这样最后的供应点集即是最终结果。这一算法运用空间邻近相关性的特性，有目的地对部分相邻或相关的候选点进行比较检验和取代，避免了很多不必要的计算，大大改善了算法的处理速度，这对大数据量的运算是非常重要的。根据 Densham 和 Rushtom 的试验，与 Teitz－Bart 算法相比，新的算法仅需要 30%～50%的处理时间。

§8.8 空间距离的量算

1. 点/点距离计算

(1)平面距离与角度。设有平面上的两点 $P_1(x_1,y_1)$与 $P_2(x_2,y_2)$，则两点之间的平面距离为：

$$|P_1P_2| = \sqrt{(x_1-x_2)^2+(y_1-y_2)^2} \qquad (8—30)$$

矢量 P_2 到 P_1 之间的夹角为：

$$sin\theta=(x_1-x_2)/|P_1P_2|, cos\theta=(y_2-y_1)/|P_1P_2|$$

(2)空间直线距离。设空间两点为 $P_1(x_1,y_1,z_1)$与 $P_2(x_2,y_2,z_2)$，则两点之间的空间直线距离为：

$$|P_1P_2|=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2} \qquad (8-31)$$

(3)球面距离。在航空与航海等情况中，作业范围较大，大圆线则是球面上两点之间的最短距离。给定球面两点 A(ψ1,λ1)和 B(ψ2,λ2)，如图 8－28 所示，圆弧长为：

$$cos(S)=sin\psi_1 \cdot sin\psi_2+cos\psi_1 \cdot cos\psi_2 \cdot cos(\lambda_2-\lambda_1)$$

$$S=arccos(sin\psi_1 \cdot sin\psi_2+cos\psi_1 \cdot cos\psi_2 \cdot cos(\lambda_2-\lambda_1))$$

两点之间的球面距离为：

$$L=R \cdot S \cdot \pi/180° \qquad (8-32)$$

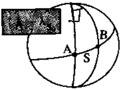


图 8—28 球面距离计算

2. 点/线距离计算

(1)点/线段最短距离。如图 8—29 所示，设线段 P_1P_2 位于横坐标轴，端点为 P_1 和 P_2。要计算点 p 到线段 P_1P_2 的距离。通过两端点 P_1 和 P_2，与线段 P_1P_2 垂直的直线将平面区域划分为 A_1、A_2、A_3。则点 p 到线段 l 的最短距离的计算方法为：

$$d_{min}=\begin{cases} d(p,P_1) & p\in A_1 \\ d(p,p^*) & p\in A_2 \\ d(p,P_2) & p\in A_3 \end{cases} \qquad (8-33)$$

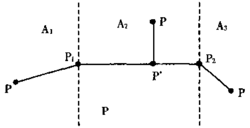


图 8—29 点/线段最短距离计算

(2)点/线段垂直距离。设有 A、B 两点组成的直线 l 为 ax＋by＋c＝0，则点 P(X,Y)到直线 l 的距离计算式为：

$$distance(P,l) = |aX+bY+c| / \sqrt{a^2+b^2} \qquad (8-34)$$

(3)点/线段平均距离。点/线段平均距离 d(p,l)等于点 p 与直线段两个端点 p_a 和 p_b 的距离值的平均值。即：

$$d(p,l) = (d(p,p_a)+d(p,p_b))/2 \qquad (8-35)$$

(4)点/线段最大距离。点/线段最大距离 $d_{max}(p,l)$等于点 p 与直线段两个端点 P_a 和 P_b 的距离中的最大者。即：

$$d_{max}(p,l) = max(d(p,p_a),d(p,p_b)) \qquad (8-36)$$

3. 点/面距离计算

(1)点/面最短距离。点 p 到面 A_n 的最短距离肯定在面的边界上，因此点面最短距离为：

$$d_{min}(p,A_n) = min\{d_{min}(p,l_i), i=1,\cdots\cdots,n\} \qquad (8-37)$$

其中 l_i 为组成面的边界线段。

(2)点/面最大距离。点/面最大距离为：

$$d_{max}(p,A_n) = max\{d_{max}(p,l_i), i=1,\cdots\cdots,n\} \qquad (8-38)$$

(3)点/面平均距离计算。点到面的平均距离计算比较复杂，AtsuyukiOkabe，HarveyH. Miller(1996)将多边形划分为三角形，按照下式计算点到面的平均距离。

$$d_{avg}(p,A_n) = \sum_{i=1}^{n-2} \frac{|T_i|}{|A_i|} d_{avg}(p,T_i) \qquad (8-39)$$

其中，$|T_i|$表示三角形的面积。$d_{avg}(p,T_i)$表示点 p 到 T_i 的平均距离。

该方法在时间复杂度上非常大，首先必须将多边形剖分为三角形，按照积分的方法计算点到每个三角形的平均距离，然后才能得出点到面的平均距离。为了高效求取点到平面的平均距离，必须在面目标内寻找一点，转换为点/点距离的计算。传统方法将点/面距离近似表示为点到质心的距离。但对于凹多边形，可在多边形内求取一内点，将点到面的距离近似表示为点到该内点的距离。

4. 线/线距离计算

(1)线/线最短、最大距离。相交线之间的最短距离为零，否则，设有两线段 l_1 和 l_2，端点分别为(p_{1s}，p_{1e})与(p_{2s}，p_{2e})。则线/线之间的最短距离和最大距离为(AtsuyukiOkabe，HarveyH. Miller，1996)：

$$d_{min}(l_1,l_2)=min\{d_{min}(p_{1s},l_2),d_{min}(p_{1e},l_2),\\ d_{min}(p_{2s},l_1),d_{min}(p_{2e},l_1)\} \quad (8—40)$$

$$d_{max}(l_1,l_2)=max\{d_{max}(p_{1s},l_2),d_{max}(p_{1e},l_2),\\ d_{max}(p_{2s},l_1),d_{max}(p_{2e},l_1)\} \quad (8—41)$$

(2)线/线平均距离计算。假设线段 l_1 位于 X 轴，$\{(x,0)|a\leqslant x\leqslant b\}$。$l_2$ 位于直线 $Y=kX$ 上，$\{(X,Y)|c\leqslant X\leqslant d,kc\leqslant Y\leqslant kd\}$。对于 l_2 上的任意一点 $P(X,Y)$，P 到 l_1 的平均距离为：

$$d_{avg}(p,l_1)=\frac{1}{b-a}\int_a^b f(x)dx$$

其中 $f(x)=\sqrt{(X-x)^2+(kX)^2}$。在 l_2 上，每间隔 dX 取点，则取样点总数 $n=(d-c)/dX$。从而线段 l_2 到 l_1 的平均距离为：

$$d_{avg}(l_2,l_1)=(d_{avg}(p_1,l_1)+d_{avg}(p_2,l_1)+\cdots\cdots+d_{avg}(p_n,l_1))/n$$

$$=\frac{1}{d-c}\int_c^d d(p,l_1)dx \quad (8—42)$$

5. 线/面距离计算

(1)线/面最短、最大距离。面由一系列有序线段构成，线/面最短、最大距离计算可转化为线/线最短与最大距离计算。求取线到组成面的每一条线段的最短与最大距离，其中的最小与最大值即为线/面之间的最短与最大距离。

(2)线/面平均距离计算。先求出面的内点，从而将线/面平均距离，近似转化为点/线平均距离计算。

§8.9 空间分析模型

1. 模型的概念和模型的生成

（1）模型（Model）的概念与特点。模型是客观世界的一个表征和体现，同时又是客观事物的抽象和概括。人们认识和研究客观世界一般有三种方法：逻辑推理法、实验法和模型法。模型法是我们了解和探索客观世界的最有力、最方便、最有效的方法。客观世界的实际系统是极其复杂的，它的属性也是多方面的。但是，建立模型决不能将所有这些因素和属性都包括进去，只能根据系统的目的和要求，抓住本质属性和因素，准确地描述系统。

一般地，模型具有以下几个特点：

· 模型比现实世界容易操作，尤其一些参数值的改变在模型中操作比在实际问题中操作更容易；

· 有些不能做实验的实际问题，可以通过建立模型来进行描述；

· 有些变量在现实中需要很长时间才能观察出它的变化情况，但用模型研究则能很快看出变化规律，从而能最迅速地抓住本质特征；

· 用模型研究变量之间的关系，可以节约时间，降低费用；

· 可以通过模型进行灵敏度分析，以便看出哪些因素对系统影响更大。

（2）模型的表示方法。模型的表示方法有两种表示形式：一阶谓词逻辑表示形式与关系表示形式。一般采用容易为用户接受的关系表示方法。它将模型看成一个虚关系，它的项由模型的输入项和输出项组成。这个虚关系是模型输入项与输出项的笛卡尔乘积的子集。这样，在创建模型时，就象创建关系数据库中的关系一样。此外，还需定义关系中各项的域值，以及输入项和输出项之间的关系。

这样，可以通过对以下几项的说明来描述某个模型：

a. 模型标识（例如模型名、模型编号）

b. 输入项

c. 输出项

d. 输入项限制（任选项）

e. 依赖关系（任选项，即输入项和输出项之间的关系）

f. 模型类型

g. 模型描述

（3）模型的分类。模型可以分为两类，即形象模型和抽象模型。前者包括直观模型、物理模型等，后者包括思维模型、符号模型、数学模型和仿真模型等。但这些模型之间是互相联系的，一个系统中往往需要多种类型的模型。

①直观模型。只供展览用的实物模型以及玩具、照片等。通常是把原型的尺寸放大或缩小。

②物理模型。主要是直观模型的进一步改进，它不仅可以显示原形的外形和特征，而且可以用来进行模拟实验，间接地研究原形外形的某些规律。

③思维模型。指通过人们对原型的反复认识，将获取的知识以经验形式直接存储于大脑中，从而可以根据思维或直觉作出相应的决策。专家系统中的专家知识就是一种思维模型。

④符号模型。是在一些约定或假设下借助于专门的符号、线条等，按一定形式组合起来描述原型。地图就是一种典型的符号模型。

⑤仿真模型。这是通过计算机上运行的程序表达的模型。物理模型和数学模型一般可以转化为仿真模型，常见的有三维仿真模型。

⑥数学模型。是用的最多、最广的一种模型。它是用字母、数字和其它数学符号构成的等式或不等式，或用图表、图像、框图、数理逻辑来描述事物的，它是真实系统的一种抽象，是研究和掌握系

统运动规律的有利工具，是分析、设计、预报、控制实际系统的基础。

(4)模型生成过程的特点。模型生成是指在系统内形成一组能反映实体运动形态的模型。一般具有以下几个特点：

·模型在系统内的主要存储方式可以是程序式或非程序式的。基于数据的模型表示和基本知识的模型表示都可以作为模型的基本形态。

·在生成模型的过程中，采用定量建模和推理分析相结合的技术。其中定量建模用于确定模型的数学形态和有关参数。

·模型生成在人机交互中完成。在传统建模过程中，人的作用和机器的作用是界限分明的，而在模型生成技术中这种分界线是模糊的。

·模型生成应该是一个动态过程。模型的生成环境发生变化时，所生成的模型也不会完全相同。

·模型生成的一般步骤。模型的描述方式和模型的生成关系密切。不同的描述方式有不同的建模步骤，下面介绍模型生成的一般步骤。

——建模准备：了解所分析的问题的实际背景，弄清建立模型的目的，掌握所分析的对象的各种信息。

——模型假设：对所研究的问题进行必要的简化，并用明确的语言对模型作出假设，即用语言对模型予以描述。

——模型建立：利用适当的数学工具建立各种量之间的关系，给出具体的数学结构。

——模型求解：求解数学方程参数，绘制图形，证明定理，进行逻辑运算等，一般利用计算机完成。

——模型分析：对上述结果进行数学上的分析，给出预测或给出最优决策或控制。

——模型检验：用实际现象、数据等检验模型的合理性和适用性。

——模型应用：如果检验结果符合要求，就可实际应用；若检验结果不符合实际情况，还需对模型进行修改、补充、假设、重新建模。

(6)模型的动态生成。传统的建模方法，在实际应用中取得的效果并不理想，主要表现在：

①建模时没有考虑环境、参数及一些外因的变化，造成所建立的模型通用性差，一般只能使用一次。

②模型和数据的接口没有设计好，往往一种模型只能用固定的一组数据。

③模型和模型之间的数据传输的接口没有设计好，只能单个模型辅助决策，模型组合能力差。

④传统的建模方法没有使建模者和利用模型决策者进行充分的交互。用户只能在已建好的模型中选择使用。

根据系统的内因和外因变化修改模型，使之跟踪系统运动形态的过程称为模型的动态生成。

模型的动态生成建立在模型生成的基础上，如果系统内还没有描述实体的模型，那么首先要生成一个模型。一般模型的动态生成包括仿真环境跟踪、模型参数更新、模型重构和调整推理等方面的内容。

· 仿真环境跟踪(Tracing Simulation Environment)。改变系统仿真环境是动态生成模型的基础工作。仿真环境发生变化，模型一般情况下总要发生变化。有时模型对环境变化的反应不够敏感，环境变化不大时模型的分析精度仍可以使用户满意，那就没有必要对模型进行修改。

· 模型参数更新。当系统的内因和外因发生变化时，模型的结构不发生变化，仅变化模型的参数就可以了。一般用人机交互的方式来修改模型的参数。

· 模型重构(Model Restructure)。系统的环境变化和内部因素的变化都会引起模型结构的变化，这就涉及到模型重构的问题。

模型的重构主要指在模型中增删一些影响因素。一般当模型参数更新的方法不能提高模型对于系统的解释能力时，应该考虑模型重构的问题。

模型重构是一个相当复杂的问题，不同的模型生成原理对应着不同的模型重构方法。重构的内容主要涉及到模型的维数、阶数、约束条件等。如果发生了多重共线问题，那么重构就要重新选择模型的基底。

・调整推理(Reasoning Adjustment)。系统的内因和外因的变化不仅影响模型的计算结果，而且还影响模型的推理分析。因此模型中的推理机制也应该适时地作调整，这种调整推理对模型重构起着重要的指导作用。但是，调整的方向、调整到什么部位、调整多大幅度都有一定的难度。

2. GIS 的空间分析模型

(1)GIS 空间分析模型的概念与特点。空间分析模型是指用于 GIS 空间分析的数学模型。

GIS 空间分析模型是在 GIS 空间数据基础上建立起来的模型，它和广义的模型概念既有区别，又相互联系，表现在：

①空间定位是空间分析模型特有的特性，构成空间分析模型的空间目标(点、弧段、网络、面域、复杂地物等)的多样性决定了空间分析模型建立的复杂性。

②空间关系也是空间分析模型的一个重要特征，空间层次关系、相邻关系以及空间目标的拓扑关系决定了空间分析模型建立的特殊性。

③包括笛卡儿坐标、高程、属性以及时序特征的空间数据极其庞大，大量的空间数据通常用图形的方式来表示，这样，由空间数据构成的空间分析模型具有可视化的图形特征。

④空间分析模型不是一个独立的模型实体，它和广义模型中抽象模型的定义是交叉的。GIS 要求完全精确地表达地理环境间复杂的空间关系，因而常使用数学模型，此外，仿真模型和符号模

型也在GIS中得到了很好的应用。

我们可以这样来认识空间分析模型的概念和意义：

——空间分析模型是联系GIS应用系统与专业领域的纽带，必须以广泛、深入的专业研究为基础；

——空间分析模型是综合利用GIS中大量数据的工具，数据的综合分析和应用主要通过模型来实现；

——空间分析模型是分析型和辅助决策型GIS区别于管理型GIS的一个重要特征，是解决空间分析和辅助决策问题的核心。

(2)空间分析模型的类型。GIS中空间分析模型的分类问题，目前研究的很少。这里，把空间分析模型分为以下几种类型：

·空间分布分析模型。用于研究地理对象的空间分布特征。主要包括：空间分布参数的描述，如分布密度和均值、分布中心、离散度等；空间分布检验，以确定分布类型；空间聚类分析，反映分布的多中心特征并确定这些中心；趋势面分析，反映现象的空间分布趋势；空间聚合与分解，反映空间对比与趋势。

·空间关系分析模型。用于研究基于地理对象的位置和属性特征的空间物体之间的关系。包括距离、方向、连通和拓扑等四种空间关系。其中，拓扑关系是研究得较多的关系；距离是内容最丰富的一种关系；连通用于描述基于视线的空间物体之间的通视性；方向反映物体的方位。

·空间相关分析模型。用于研究物体位置和属性集成下的关系，尤其是物体群(类)之间的关系。在这方面，目前研究得最多的是空间统计学范畴的问题。统计上的空间相关、覆盖分析就是考虑物体类之间相关关系的分析。

·预测、评价与决策模型。用于研究地理对象的动态发展，根据过去和现在推断未来，根据已知推测未知，运用科学知识和手段来估计地理对象的未来发展趋势，并作出判断与评价，形成决策方案，用以指导行动，以获得尽可能好的实践效果。

3. GIS 中常用的空间分析模型

（1）相关分析模型。GIS 地理数据库中存储的各种自然和人文地理要素（现象）的数据并不是孤立的，它们相互影响、相互制约，彼此之间存在着一定的联系。相关分析模型就是用来分析研究各种地理要素数据之间相互关系的一种有效手段。

地理数据库中各种地理要素数据之间的相关关系，通常可以分为参数相关和非参数相关两大类。其中，参数相关又可分为简单（两要素）线性相关、多要素间的复相关、偏相关和典型相关；非参数相关可以分为顺序（等级）相关和二元分类相关。这里，只介绍简单线性相关、偏相关和复相关分析模型。其他相关模型读者可参阅有关文献。

——简单线性相关模型。在一般情况下，当两种要素之间为线性相关时，就要研究它们之间的相关程度和相关方向。所谓相关程度，指它们之间的相关关系是否密切；所谓相关方向，就是两种要素之间相关的正负。相关程度和相关方向，可以用相关系数来衡量。

设 X 和 Y 为两种地理要素（现象），X_j 和 Y_j 分别为它们的样本统计值（j=1，2，…，n），则它们之间的相关系数模型为：

$$\gamma=\frac{\sum_{j=1}^{n}(X_j-\overline{X})(Y_j-\overline{Y})}{\sqrt{\left[\sum_{j=1}^{n}(X_j-\overline{X})^2\cdot\sum_{j=1}^{n}(Y_j-\overline{Y})^2\right]}}=\frac{\sigma_{xy}}{\sqrt{\sigma_x^2\cdot\sigma_y^2}} \quad (8—43)$$

式中：

$$\overline{X}=\frac{1}{n}\sum_{j=1}^{n}X_j,\overline{Y}=\frac{1}{n}\sum_{j=1}^{n}Y_j,$$

$$\sigma_x^2=\sum_{j=1}^{n}(X_j-\overline{X})^2,$$

$$\sigma_y^2=\sum_{j=1}^{n}(Y_j-\overline{Y})^2,$$

$$\sigma_{xy}=\sum_{j=1}^{n}(X_j-\overline{X})(Y_j-\overline{Y})$$

相关系数的取值范围为$-1\leqslant\gamma\leqslant+1$。当相关系数为正时，表示两种要素之间为正相关；反之，为负相关。相关系数的绝对值$|\gamma|$越大，表示两种要素之间的相关程度越密切，$\gamma=+1$为完全正相关，$\gamma=-1$为完全负相关，$\gamma=0$为完全线性无关。

——多要素相关模型。

・任意两种要素间的相关系数模型。设有一组地理要素变量$X_1,X_2,\cdots,X_m$，统计n个样本，则n个样本m个指标可构成一个$n\times m$阶的原始数据矩阵。此时，任意两种要素间的相关系数模型为：

$$\gamma_{ik}=\frac{\sum_{j=1}^{n}(X_{kj}-\overline{X}_i)(X_{kj}-\overline{X}_j)}{\sqrt{[\sum_{j=1}^{n}(X_{kj}-\overline{X}_i)^2\cdot\sum_{j=1}^{n}(X_{kj}-\overline{X}_j)^2]}}=\frac{\sigma_{ik}}{\sqrt{\sigma_i^2\cdot\sigma_k^2}} \qquad (8—44)$$

式中$\sigma_{ik},\sigma_k^2,\sigma_i^2$分别为样本的协方差和方差。

・偏相关系数模型。当研究某一种要素对另一种要素的影响或相关程度，而把其他要素的影响完全排除在外，单独研究那两种要素之间的相关系数时，就要使用偏相关分析方法，偏相关程度用偏相关系数来衡量。

若i,j,k代表变量$\{x_1,x_1,\cdots,x_m\}$中任意三种不同的变量，则所有一阶偏相关系数模型如下：

$$\gamma_{ij\cdot k}=\frac{\gamma_{ij}-\gamma_{jk}\cdot\gamma_{ik}}{\sqrt{[(1-\gamma_{jk}^2)(1-\gamma_{ik}^2)]}} \qquad (8-45)$$

式中：γ_{ij}、γ_{ik}、γ_{jk}为单相关系数。

逐次使用递归公式

$$\gamma_{ij\cdot ck}=\frac{\gamma_{ij\cdot c}-\gamma_{jk\cdot c}\cdot\gamma_{ik\cdot c}}{\sqrt{[(1-\gamma_{jk\cdot c}^2)(1-\gamma_{ik\cdot c}^2)]}} \qquad (8-46)$$

就可以得到任意阶的偏相关系数。其中,c 是其余变量的任意子集合。

・复相关系数模型。以上都是在把其他要素的影响完全排除在外的情况下研究两种要素之间的相关关系。但是实际上,GIS 的空间分析中,一种要素的变化往往要受到多种要素的综合影响,这时就需要采用复相关分析方法。所谓复相关,就是研究几种地理要素同时与某一种要素之间的相关关系,度量复相关程度的指标是复相关系数。

设因变量为 Y,自变量为 $X_1, X_1, \cdots, X_k$,则 Y 与 $X_1, X_2, \cdots, X_k$ 的复相关系数计算公式为:

$$R_{Y\cdot 1,2,\cdots,k}=\sqrt{[1-(1-\gamma_{Y\cdot 1}^2)(1-\gamma_{Y\cdot 2,1}^2)\cdots(1-\gamma_{Y\cdot k,1,2,\cdots,k-1}^2)]} \qquad (8-47)$$

作为特例,三个变量(Y, X_1, X_2)之间的复相关系数的计算公式为:

$$R_{Y\cdot 1,2}=\sqrt{\frac{\gamma_{Y1}^2+\gamma_{Y2}^2-2\gamma_{Y1}\gamma_{12}\gamma_{Y2}}{1-\gamma_{12}^2}} \qquad (8-48)$$

(2)趋势面分析模型。在 GIS 的空间分析中,经常要研究某种现象的空间分布特征与变化规律。许多现象在空间都具有复杂的分布特征,它们常常呈现为不规则的曲面。欲研究这些现象的空间分布趋势,就要用适当的数学方法将现象的空间分布及其区域变化趋势模拟出来,这就是趋势面分析方法。

趋势面分析,是用一个多项式对地理现象的空间分布特征进行分析,用该多项式所代表的曲面来逼近(或拟合)现象分布特征的趋势变化,也就是用数学方法把观测值分解为两个部分:趋势部分和偏差部分。趋势部分反映区域性的总的变化,受大范围的系统

性因素的控制；偏差部分反映局部范围的变化特点，受局部因素和随机因素的控制。

——基本原理。设 $Z_j(x_j, y_j)$ 表示所分析现象的特征值，即观测值。趋势面分析就是把观测值 Z 的变化分解成两个部分，即：

$$Z_j(x_j, y_j) = f(x_j, y_j) + \sigma_j \qquad (8—49)$$

式中：$f(x_j, y_j)$ 为趋势值，σ_j 为剩余值。

我们可以用回归方法求得趋势值和剩余值，即根据已知数据 Z 的一个回归方程 f(x,y)，使得

$$Q = \sum_{j=1}^{n} [Z_j - f(x_j, y_j)]^2 \qquad (8—50)$$

达到极小。这实际上是在最小二乘法意义下的曲面拟合问题，即根据观测值 Zj(xj,yj) 用回归分析方法求得一个回归曲面

$$\hat{Z} = f(x, y) \qquad (8—51)$$

而以对应于回归曲面的值

$$\hat{Z}_j = f(x_j, y_j)$$

作为趋势值，以残差

$$Z_j - \hat{Z}_j$$

作为剩余值。

——多项式趋势面的数学模型。在趋势面分析中，我们通常选择多项式作为回归方程，因为任何一个函数在一个适当的范围内总是可以用多项式来逼近，而且调整多项式的次数可以使求得的回归方程适合问题的需要。

当某一地理现象的特征值在空间的分布为平面、二次曲面即抛物曲面、三次曲面、四次曲面、五次曲面或六次曲面时，可分别用一次多项式、二次多项式、三次多项式、四次多项式、五次多项式或六次多项式来拟合。多项式数学模型中各项的排列顺序有一定规律，便于编程计算。

——多项式趋势面数学模型的解算实际上是求多项式系数的最佳无偏估值问题。最小二乘法可以给出多项式系数的最佳线性

无偏估值，这些估值使残差平方和达到最小。所以求回归方程也就是要求根据观测值 $Z_j(x_j, y_j)(j=1,2,\cdots,n)$，确定多项式的系数 $a_0, a_1, \cdots$，以使残差平方和最小，即：

$$Q=\sum_{j=1}^{n}(Z_j-\hat{Z}_j)^2=\min \tag{8—52}$$

我们记 $x=x_1, y=x_2, x^2=x_3, xy=x_4, y^2=x_5, \cdots\cdots\cdots$，则多项式可以写为：

$$\hat{Z}=a_0+a_1x_1+a_2x_2+a_3x_3+\cdots+a_px_p$$

这样，多项式回归问题就可以转化为多元线性回归问题来解决。现在，残差就是：

$$Q=\sum_{j=1}^{n}[Z_j-(a_0+a_1x_1+a_2x_2+a_3x_3+\cdots+a_px_p)]^2$$

根据最小二乘法原理，我们要选择这样的系数 $a_0, a_1, \cdots\cdots a_p$ $(p<n)$，以使 Q 达到极小。为此，求 Q 对 $a_0, a_1, \cdots\cdots, a_p$ 的偏导数，并令其等于零，则得正规方程组。解此正规方程组，即得 $p+1$ 个系数 $a_0, a_1, \cdots, a_p$。

在原始数据量很大的情况下，用矩阵方法求解在计算机上实现是困难的，因为占据存储空间太大。所以，一般采用高斯主元消去法或正交变换法求解正规方程组。

——趋势面拟合程度的检验。趋势面的拟合程度就是趋势面对原始数据面的逼近度。这里介绍两种检验方法：

·F—分布检验

检验统计量为：

$$F=\frac{U/P}{Q/(n-P-1)} \tag{8—53}$$

式中：U 为回归平方和，Q 为剩余平方和，P 为多项式的项数（不含常数项），n 为观测点数。在给定置信水平 α 的条件下，若 $F>F_\alpha$，则趋势面拟合效果显著，否则不显著。

·拟合指数公式检验

拟合指数公式为：

$$C=\left[1-\frac{\sum(Z_j-\hat{Z}_j)^2}{\sum(Z_j-\overline{Z})^2}\right]\times100\% \qquad (8\text{—}54)$$

式中：C 为拟合指数，Z_j 为第 j 点的观测值，$\hat{Z}_j$ 为第 j 点的趋势值，$\overline{Z}$ 为全部观测值的平均值。当 C＝100％时，表明趋势值在所有观测点上都与实际值吻合，但这种情况是很少的。当 C＝75％以上时，拟合误差均在 10％以下，这时可以认为趋势面的拟合效果良好。

(3)预测模型。GIS 地理数据库的数据除了反映各种自然和人文要素(现象)的空间分布特征和相互关系外，还能反映地理要素的动态发展规律，并用于预测分析。这种预测分析是建立在现象间因果关系的基础上的，即某些现象作为原因，另一种现象作为结果，原因与结果的关系可以用确定的函数来描述，函数中的参数能说明这种因果关系的本质。预测模型常用于判断结果随原因的变化而变化的方向和程度，用于推断随时间发生变化的大小。

所谓回归模型方法，就是从一组地理要素(现象)的数据出发，确定这些要素数据之间的定量表述形式，即建立回归模型。通过回归模型，根据一个或几个地理要素数据来预测另一个要素的值。这种回归模型就是一种预测模型。

——一元回归模型。一元回归模型表示一种地理要素(现象)与另一种地理要素之间的依存关系，另一种要素作为它的分布与发展的最重要的原因。模拟一元回归模型时，必要条件是具有两相应的变量系列，其中同一系列的每个元素完全相应于另一序列的元素，这时可以实现内插和外推两个任务。

我们用多项式方程作为一元回归的基本模型：

$$Y=a_0+a_1x+a_2x^2+a_3x^3+\cdots\cdots a_mx^m+\varepsilon$$

式中：Y 为因变量，X 为自变量，$a_0,a_1,\cdots,a_m$ 为回归系数，ε 为剩余误差。

上式中多项式的次数由地理要素之间的关系确定。通常是采用函数逼近的方法来确定多项式的次数，首先从一次多项式开始，直至多项式的剩余误差平方和小于某个给定的任意小数为止。

利用多项式进行预测，最主要的问题是求解方程式的系数 a_0，$a_1, \cdots, a_m$。通常采用最小二乘法求解。求得系数后，就可以用这些系数来解决内插和外推的问题。

回归模型的精度，通常可通过求 ε 来确定。根据多项式有：

$$\varepsilon_j = Y_j - (a_0 + a_1 x_j + \cdots + a_m x_j^m) = Y_j - \hat{Y}_j \qquad (8—55)$$

式中：$\hat{Y}_j$ 为计算值。

根据最小二乘法原理，ε_j 的平方和为最小是最好的，一般是采用回归方程的剩余标准差来估计，即

$$S = \sqrt{\frac{1}{n-2}\sum_{j=1}^{n}(Y_j - \hat{Y}_j)^2} \qquad (8—56)$$

S 的大小反映回归模型的效果。

关于回归效果的显著性检验，可以证明它是一个具有自由度 $(1, m-2)$ 的 F 变量，即

$$F_{(1,m-2)} = \frac{\gamma^2}{1-\gamma^2}(m-2) \qquad (8—57)$$

式中：γ 为相关系数。

可见，一元回归时，回归效果的好坏可以通过相关系数的检验来鉴别。

——多元线性回归模型。多元线性回归模型表示一种地理现象与另外多种地理现象的依存关系，这时另外多种地理现象共同对一种地理现象产生影响，作为影响其分布与发展的重要因素。

设变量 Y 与变量 $X_1, X_2, \cdots, X_m$ 存在着线性回归关系，它的 n 个样本观测值为 $Y_j, X_{j1}, X_{j2}, \cdots X_{jm}(j=1,2,n)$，于是多元线性回归的数学模型可以写为：

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1m} \\ 1 & X_{21} & X_{22} & \cdots & X_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} \quad (8-58)$$

可采用最小二乘法对上式中的待估回归系数 $\beta_0,\beta_1,\cdots,\beta_m$ 进行估计，求得 β 值后，即可利用多元线性回归模型进行预测了。

计算了多元线性回归方程之后，为了将它用于解决实际预测问题，还必须进行数学检验。多元线性回归分析的数学检验，包括回归方程和回归系数的显著性检验。

回归方程的显著性检验，采用统计量：

$$F=\frac{U/m}{Q/(n-m-1)} \quad (8-59)$$

式中：$U=\sum_{j=1}^{n}(\hat{Y}_j-\overline{Y})^2$，为回归平方和，其自由度为 m；$Q=\sum_{j=1}^{n}(Y_j-\hat{Y})^2$，为剩余平方和，其自由度为(n－m－1)。

利用上式计算出 F 值后，再利用 F 分布表进行检验。给定显著性水平 α，在 F 分布表中查出自由度为 m 和(n－m－1)的值 F_α，如果 $F \geqslant F_\alpha$，则说明 Y 与 $X_1,X_2,\cdots,X_m$ 的线性相关密切；反之，则说明两者线性关系不密切。

回归系数的显著性检验，采用统计量：

$$F=\frac{(b_i-\beta_i)^2/C_{ii}}{Q/(n-m-1)} \quad (8-60)$$

式中，C_{ii}为相关矩阵 $C=A^{-1}$的对角线上的元素。

对于给定的置信水平 α，查 F 分布表得 $F_\alpha(n-m-1)$，若计算值 $F_i \geqslant F_\alpha$，则拒绝原假设，即认为 X_i 是重要变量，反之，则认为 X_i 变量可以剔除。

多元线性回归模型的精度，可以利用剩余标准差

$$S=\sqrt{Q/(n-m-1)} \quad (8-61)$$

来衡量。S越小，则用回归方程预测Y越精确；反之亦然。

(4)聚类模型。

这里只介绍模糊聚类模型。设有m个变量，每个变量有n个样本，其数据矩阵为：

$$X=(X_{ij})_{n\times m} \tag{8—62}$$

对n个样本进行模糊聚类的步骤如下：

·统计指标的数据标准化。可采用标准差标准化公式，也可采用极差标准化公式。若采用前者，则计算结果尚需压缩到[0,1]区间；若采用后者，则计算结果已压缩到[0,1]区间。

·标定。即计算出衡量样本(被分类对象)间相似性程度的统计量γ_{ij}，建立论域$U=\{u_1,u_2,\cdots,u_n\}$上的相关关系$\tilde{R}$，$\mu_R=(u_i,u_j)$表示u_i与u_j按分类特性的相似性程度，$\tilde{R}$可用相似矩阵R表示。γ_{ij}可以计算得出，也可采取专家评分的方式给出。

·将模糊关系矩阵改造成模糊等价关系矩阵$\tilde{R}\rightarrow R^*$

采用求传递背包的平方法：

$$R\rightarrow R^2\rightarrow R^4\rightarrow\cdots\rightarrow R^{2k}=R^*$$

$$2^{k-1}<n<2^k$$

即

$$k-1<log2^n\leqslant k$$

·聚类。模糊等价关系矩阵R^*的元素表示被分类对象彼此之间的相似程度，把R^*的元素从大到小排列作为规定的λ水平值$(0\leqslant\lambda\leqslant1)$，使

$$CR_\lambda(i,j)=\begin{cases}0,若\ R^*(i,j)\geqslant\lambda\\1,若\ R^*(i,j)<\lambda\end{cases}$$

利用求得的模糊等价关系的λ水平截集进行分类，分类由粗到细。选R^*元素的最小值为λ值，则分为一类；选第二个最小值为λ值分为两类；选第K个最小值为λ值就分为K类；……。若λ=1，则各个样本自成一类。

4. **模型库及其管理**

(1)模型库的基本概念和表示形式。

①模型库的基本概念。随着GIS中空间分析特别是空间决策支持技术的进步，基于模型的决策支持已经由单模型辅助决策发展到多模型辅助决策，模型的组织形式也由模型软件包(简称模型包)发展到模型库。

模型库是在计算机中按一定的组织结构形式存储多个模型的集合体，在模型库管理系统下得到有效的管理。同地理数据库与地理数据库管理系统相结合构成地理数据库系统一样，模型库与模型库管理系统相结合构成模型库系统。当然，模型库与数据库在本质上是有很大差别的。

模型库不同于模型包。模型包虽然也含有多个模型，但各模型之间没有任何联系，也没有任何约束，即不是通过一定组织结构的存储形式把模型组织起来的。在模型库中，模型是按照一定的组织形式存储起来的，这种组织结构形式有利于对模型进行有效管理和使用，提高了多模型的组合能力，有利于提高模型组合的辅助决策作用。

②模型库的表示形式。数学模型的一般表示形式是方程式，它反映了模型中变量之间的关系、约束条件及其目标。方程形式不适合计算机求解，在计算机中一般是将求解方程的算法编为程序形式，包括源程序(用算法语言描述)和目标程序(经过编译形成机器语言程序)。所以，在计算机中，模型是以程序形式表示的。

(2)模型库的组织和存储。模型库由模型字典库和模型文件库组成。

①模型字典库。模型字典库的内容包括模型的编号、名称与模型文件等的说明。

模型字典是模型文件的索引，便于模型与模型文件的联系，便于对模型进行分类(即对模型字典进行分类)，便于对模型和模型文件的查询和修改(对模型的算法、参数以及有关模型说明的修改)，包括增加、删除和更新。

模型字典库一般以采用数据库形式的组织结构为宜，即按照

关系数据库的组织形式存放模型字典的内容。按照模型分类可分别建立各类模型字典库，一个库存放一类模型，每个模型是一个记录，每个记录由模型的编号、名称、文件名等数据项组成。这种组织存储形式便于模型的分类、查询和修改。

②模型文件库。模型文件是模型的主体，一个模型至少有 2～4 个模型文件，其中源程序文件和目标程序文件是主要的模型文件，此外还有模型的说明文件和数据描述文件。

模型文件通常以建立子目录的方法存储，即把模型文件都建立在子目录下。子目录的建立可采用两种形式：一是按模型分类建立子目录，每类模型建立一个子目录，该类中模型的所有模型文件都存储在此子目录下，模型文件库和模型字典库一一对应；二是按模型文件的类别建立子目录，建立 2～4 个子目录，分别存放各类模型文件，即所有模型的源程序文件存放在一个子目录下，所有模型的目标程序文件存放在另一个子目录下。

模型文件特别是目标程序文件的调用即模型的运行，与模型文件的存储方式有直接关系。一般是通过模型字典库，沿着模型文件的存储路径，找到具体的模型文件，然后利用运行某模型文件的命令启动运行。

(3)模型库管理系统。模型库管理技术是适应 GIS 的空间分析和空间决策支持的需要而发展起来的，经历了程序文件、模型软件包和模型库管理系统三个发展阶段。在程序文件阶段，模型算法以程序文件形式存放在计算机的存储器中，由操作系统的文件管理系统统一管理；在模型软件包阶段，预先编制好的模型程序统一挂在交互菜单下，模型之间相对独立，在人机交互菜单的驱动下调用所需模型；在模型库管理系统阶段，模型不是挂在菜单下，而是在管理系统下，有效地对模型进行存储、修改、查询、调用，模型之间不仅相互独立，而且可以互相组合。

模型库管理系统的主要功能包括模型的存储管理、运行管理和建模技术等三个方面。

①模型的存储管理功能。包括模型的表示、模型的存储组织结构和模型的查询与维护。

数学模型在计算机中都是以程序形式表示的，调用数据后，执行程序就能得到结果。程序在计算机中是以文件形式存储的，称为程序文件。

模型的存储组织结构，由两部分构成，即：一部分是模型字典库，它类似于数据库的组织结构形式，但存储的不是数据字典，而是模型字典；另一部分是模型文件库，它是模型的主体，具有文件形式，按文件存储方式存储。在模型字典库中应指明模型文件的存储路径，便于查找。

模型的查询与维护，是模型库管理系统的重要功能。根据模型的存储组织结构形式，要查询模型，首先要查询模型字典库，找到需要的模型目录，再沿着模型文件的存取路径查到相应的模型文件。所以，模型的查询过程包括两部分内容：一个是模型字典库的查询，类似于数据库的查询；另一个是模型文件的查询，类似于操作系统的文件查询。模型的维护类似于数据库的维护，包括对模型进行增加、插入、删除、修改等操作，这些操作要按模型的存储组织结构形式进行。当增加、插入、删除模型时，首先要在模型字典库中增加、插入、删除模型目录，然后沿着模型存取路径在模型文件库中增加、插入、删除模型文件。当要对模型进行修改时，不修改模型目录，只需修改模型文件。

②模型的运行管理功能。包括模型程序的输入和编译、模型的运行控制、模型对数据的存取。

模型程序的输入通过编辑系统来实现。编辑系统具有输入、修改、插入模型程序的功能。用户输入的模型程序是源程序，必须通过相应语言的编译系统将其编译成目标程序，计算机才能运行。

模型程序在计算机中是这样运行的：按照模型的组织存储结构，首先在模型字典库中找到该模型目录。再按模型文件的存取路径找到模型的目标程序文件。目标程序可以在操作命令系统下独

立运行，也可以在主控程序中运行。

模型运行是需要数据的。在 GIS 中，模型库和地理数据库是相互独立的，从模型程序中存取地理数据库中的数据要通过模型库和地理数据库的接口来实现。

③模型的组合功能。计算机建模，是 GIS 的发展方向，技术还不成熟。但模型库管理系统必须具有模型的组合功能。

模型组合包括模型间的组合及模型间数据的共享和传递两个方面的问题。模型间的组合通过程序设计中的顺序结构、选择结构和循环结构等三种组织结构方式来实现。这三种组织结构方式可以嵌套使用。模型间数据的共享和传递是与模型组合相适应的，在 GIS 中模型库与地理数据库互相独立，在所有数据都存储在地理数据库中的情况下，只要解决好模型存取地理数据库数据的接口问题，就能实现模型间数据的共享和传递。

第九章　GIS 中的地图制图

空间数据经 GIS 分析和处理后，所得到的结果必须以各种形式输出给用户，供生产、研究和管理部门使用。图形、图像、图表及文本是其常用输出形式，各种类型的专题图已成为 GIS 的主要输出产品。为了能准确、清晰、高效地制作出高质量的图件，制图功能已成为 GIS 中不可缺少的重要功能之一。通常，不同的 GIS 所具备的制图功能的强弱各不相同，这已经成为评价 GIS 功能强弱的重要标志之一，该功能的强弱也直接影响它在市场上的占有率。

§9.1　GIS 中地图制图的基本原理

地图可以认为由地图图形和地图注记两部分组成。地图图形是指地图上用来表示自然和人文现象的各种地图符号，通常分为点、线、面三种形式，而地图注记是指对地图符号和地图内容的说明信息。

1. GIS 的地图制图流程和产品

(1)GIS 中地图制图的流程。在 GIS 中，地理数据库存储着地图、航空像片、遥感图像、统计资料、文本信息等多种数据，GIS 中的地图制图系统的功能就是利用这些数据制作地图。GIS 所制作的地图大多是专题地图。GIS 中制作地图的流程可由图 9—1 所示。

无论制作什么地图，首先必须根据地图用途确定地图内容，即必须确定地图上应表示哪些要素(如河流、道路等)。对于专题地图来说，地图内容可分为底图内容和专题内容。在 GIS 的地图制图

系统中,可以用人机交互的方式确定地图内容。如果是对 GIS 当前的分析结果制图时,当前的分析内容就是所需要表示的地图内容,当然也可以叠加上一些底图要素。

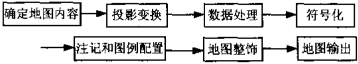


图 9—1 地图制图的流程

数据库中的空间数据或者是以地理坐标存储的,或者是以某种地图投影存储的。因此,往往需要对空间数据进行投影变换,以符合当前所需制作地图的要求。

数据处理在这里是指属性数据的处理(如分类和分级)和几何数据的处理(如坡度的计算等),是为地图内容的符号化所作的准备工作。

地图内容的符号化根据其所处理的数据类型和符号化的方法可分为对属性数据的符号化和对几何数据的符号化。对属性数据的符号化(如对道路等级、人口数的符号化)可根据数据的数学属性选择相应的地图符号视觉变量对其进行表示,而几何数据的符号化很难用视觉变量原理来解释(如对地形、气压的符号化),需要用特殊的方法(如等值线、三维立体图等)来表示。

注记和图例是地图不可缺少的,直接影响到用户对地图的理解。

要绘制一副完整的地图往往需要有图廓、图名、接图表、指北针、比例尺等各种图饰,这就是地图的整饰工作。

地图输出的主要任务是将地图数据处理和符号化的结果,变成输出装置可识别的信号,以驱动输出装置产生地图图形和地图注记。

(2)GIS 的地图产品。GIS 中的地图制图系统所产生的地图产

品可以采用矢量绘图机、栅格绘图机、打印机、图形显示器、缩微系统、磁盘等硬件输出，因而其地图产品可以是纸质地图、屏幕地图（电子地图）、数字地图等。电子地图比纸质地图具有更多的优点：制作灵活，形式极其多样，修改制作方便，制作周期短，色彩丰富，动态性强，查询方便、快捷。数字地图是以数字形式贮存在磁带、磁盘、光盘等介质上的地图，具有重要的应用价值，是GIS的重要产品。

GIS的地图产品还可按其状态分为静态地图(static map)和动态地图(dynamic map)。静态地图是指地图的幅面和内容都不变的地图。传统的纸质地图就是静态地图，尽管它有表示动态现象的成功方法（如动线法等），但它本质上是静止的。动态地图可以直观而又逼真地显示地理实体运动变化的规律和特点。如用动态地图表示人口的迁移和变化等。

2. 地图语言与符号系统

地图是空间信息可视化的一种最重要的形式，它是环境空间中地理实体集合的数量、质量、时间和空间特征全面抽象后的图形符号模型。也可以说地图是图形符号的空间集合，图形符号是地图的语言。统一、协调、美观的符号系统，体现了地图的艺术水平。

(1)地图语言。地图是反映各种自然、社会现象的质量和数量分布及其联系和变化的时空直观模型。这个定义是从外部的功能角度来下的，但是进一步还要从内部结构角度来认识地图的本质。制图者通过对客观事物的认识（观察、解译、分析、概括），建立起对于制图区域整体的空间总认识（即建立起现象与现象之间的逻辑关系）；这反映了环境信息进入制图者大脑中进行信息处理、形成概念、构筑抽象思维模型的过程。这个过程对于完整的制图过程来说还只是一部分，制图者还要借助于地图符号语言对抽象思维模型进行图形模拟，把自己对空间环境的认识通过地图的形式传输给用图者。因此，制图过程可以视为抽象思维向形象思维转化的过程。

地图符号给予地图极大的影响，它决定着地图内容能否充分表达，决定着地图是否便于阅读和使用，决定着地图信息的传输效果。地图之所以能够表示各种复杂的自然或社会现象，一个主要的原因是它拥有一套完整的、科学的地图符号系统，与风景画和像片等有着截然不同的区别。因此，地图符号系统成为现代地图学最基本的特征之一。

制图人员在进行地图制图表示时使用的是符号系统，对于这种表示，既有发送者也有接收者。自然语言是在发送者与接收者之间通过口头或书面语言传递情况、思想和感情的典型符号系统。同样，制图工作者通过视觉符号系统传递思想，这就是地图语言。最初把地图语言作为一种概念可在 M. K. Bocharov (1966)，J. Bertin(1967)等人的著作中看到。地图语言的研究是以(自然)语言的定义为基础的，作为思想的工具、通讯的手段及存储知识的方式，地图语言是一套符号系统。今天，地图的支撑媒体正在从纸张变为屏幕，而且人们开始使用多媒体地图和人机动态交互系统。在这个环境里越来越需要建立地图语言的语法、语义和语用学的理论。

为了揭开地图内部语言结构的奥秘，也许可以从下面这些联想中得到启发：建筑是建筑师用一定的建筑规则对于建筑材料的组织；棋谱是棋手用一定的棋法规则对于棋子的组织；乐曲的好坏取决于作曲家按照怎样的规则把音阶符号填人五线谱的哪些位置。语言学的知识告诉我们：词汇是语言的建筑材料，语法是组词造句的规则，它们是语言构成不可缺少的因素。只有词汇不能表达思想，还必须按照一定的方式和规则组织起来，才能成为可理解的语句，才能传达语义。因此，要从两个方面研究地图符号系统的内部结构：

· 符号变量本身的属性及其独立的功能。符号变量相当于语言学中的词汇，它是作地图上最小的图解单元。

· 符号变量与符号变量、符号层面与符号层而之间相互联系

和组织的方法及其组合功能，这相当于语言学中的语法。

(2)符号变量的种类及视觉感受效果。

①符号变量的种类。符号变量通常是指可以引起视觉差别的最基本的图形和色彩因素的变化。符号变量的研究，作为地图符号设计的基础，在提高地图符号构图规律和加强地图表示事物的效果上起了不可忽视的作用。最早对符号变量进行研究，并引入到地图学来的是法国学者 J. Bertin，他提出的符号变量是：形状、尺寸、色彩、方向、亮度和纹理。J. Bertin 的视觉变量理论引起了地图学家的广泛兴趣，许多学者提出了自己的见解。Robinson 提出的符号变量是：色相、亮度、尺寸、形状、间距、方向和位置；萨里谢夫提出了地图学中区别符号特征的绘图方法：形状、尺寸、方向、亮度、色彩和内部结构；Morrison 提出了构成地图符号的八个符号维数：尺寸、形状、色相、彩度、亮度、图案排列、图案方向和图案纹理；Board 用制图字母的概念代替符号变量概念，提出的制图字母是：形状、尺寸、图案(方向、排列)、纹理和色彩(色相、彩度、亮度)。

通过对各种符号变量分类的研究，可以认为地图符号是由下列六种符号变量构成的：颜色、形状、尺寸、方向、亮度和密度(图9—2 表示后五种符号变量在点、线、面符号中的具体体现)。

②符号变量的视觉感受效果。各种符号变量能引起视觉感受的多种效果，这些效果可归纳为七种，即整体感、差异感、等级感、数量感、质量感、动态感和立体感。

· 整体感：由不同的符号组成一个图形时，看起来是一个整体。形状、近似色、方向等可以产生较好的整体感，而亮度、尺寸、反差大的色彩等变量由于本身差别过大，整体感的效果不好。

· 差异感：凡整体感效果好的，差异感就差，反之亦然。我们常常要把某一要素的符号突出显示在图面上，即明显区别于周围环境，而利用差异大的符号变量如对比强烈的色相、亮度、密度、尺寸等有较强的差异感。

· 等级感：是指将观察对象迅速而明显地分出几个等级的效

果。居民地符号的大小，公路符号的宽窄，注记的字大，都首先反映了不同的等级。能产生明显等级感的是尺寸、亮度、密度等，而色相、形状、方向等则无明显的等级感。

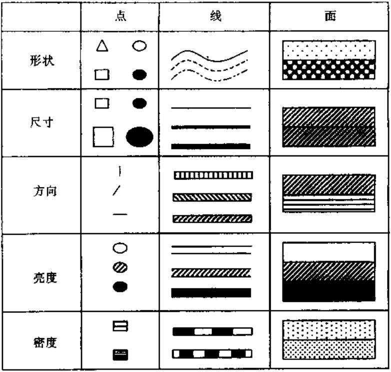


图 9-2　符号变量

・数量感：指从地图上获得各要素绝对差值的效果。尺寸变量是产生数量感的最有效的符号变量，但受图形复杂程度的影响，图形复杂时数量差别的准确性就会下降。欲建立图形符号的数量感，可以使用基本的几何图形，使问题简单化。例如，可以使用同宽度的柱状符号，以其高度的差别来模拟数量差。

・质量感：指将观察对象区分为若干不同质的几个类别的效果。它和差异感有些相似，即要明显区分出对象。但不同的是，质

量感不是单纯地把某一要素的符号突出于图面上，而是要把全部要素符号彼此明显地区别开来。色相的不同、形状的变化是产生质量差别的常用变量。

·动态感：指的是从构图上产生一种运动的视觉效果。单一的视觉变量并不能产生运动感，但有些视觉变量的有节奏的排列会产生运动感，如同一形状的尺寸有节奏改变的符号，会产生运动感。箭头符号是反映动态感的有效方法，是视觉变量的特例。

·立体感：指从两维空间的平面上产生三维空间的立体视觉效应。产生立体感较强的变量是尺寸和色彩。

通过对以上的分析，我们认为：各个相对独立的符号变量是对图上大量点状、线状和面状符号视觉差异的抽象，它们是依附于这些符号的基本属性，一般来说它们并不具有直接构图的能力。这些符号变量一般并不能相当于词的作用，面只是构词的成份，各个点、线、面符号才可以相当于词和词汇，它们才是组织句子，并用以表示一定语义的独立运用单位。

(3)符号的语言结构。语法是在语言词汇的基础上，进行抽象和概括，得出一般规律。地图语言的语法研究的是地图符号系统的构成，以及各种地图表示手段的运用和组合的规则。地图语言的语法研究，对进一步认识地图的本质、地图符号规范化和扩展地图的应用，以及促进地图生产的革新等都有着重要的意义。下面对地图符号之间的相互关系在语法层次上进行探讨。

差异与统一的关系：地图符号设计的根本问题是认识制图现象的差异和统一，分清主次。使用符号变量的数目、变化幅度及其组合程度是造成地图图面差异感和统一感的主要原因。

等级与层次的关系：地图是一个符号等级系统，符号的构成要体现地图内容的分类分级原则。

简单与复杂的关系：对于地图上表现为单一性质或单一质量变化的现象，只需调用一两个符号变量用于地图设计即可，而对于地图上多重性质或多种量值的现象，需要调用多个符号变量才能

表示明白。

空间与平面的关系:地图是以二维平面来表示三维空间的,通常采用化体为面、化面为线和化线为点的方法简化表示。

确切与概括的关系:地图表示法还能反映出地图资料的不同精度,有时在空间位置上模糊和概略的表示反而更加符合客观情况。因此,对于不同精度的制图资料,其表示方法也应与之相对应。

随意与约定的关系:地图符号的含义是随意约定的,但是一旦约定俗成以后就不能随意改变。

(4)符号系统设计的基本原理。

①地图符号的功能与本质。地图符号是用来表示地图上各种空间对象的图形记号,在某些情况下还包括与之配合使用的注记。地图符号对表达空间地理信息具有重要的作用。它能保证空间对象的位置具有较高的几何精度,从而提供了地图的可量测性。地图符号不仅能表达地面物体的存在和分布,而且可以表示地物的数量特征和质量特征,同时还可以表示较小的物体和肉眼看不见的地理现象。特别是采用地图符号后,可以使地物主次分明,使得地图所表达的地面现象主次分明,清晰易读。

地图符号的本质可以从地图符号的约定性和等价性以及地图符号的"内涵与外延"等方面来分析。所谓地图符号的约定性,是指地图符号同它所代替的抽象概念之间的一一对应关系。制图者为了传递思想和概念,采用一些图形来代替一些概念,这就是地图符号与之所代表概念之间的约定过程,从而使地图符号具有约定性。任何符号都是在社会上被一定的社会集团或科学团体所承认和共同遵守的,因而具有法定性、规定性。地图符号中,尤其是普通地图的符号,大都经过长时间的考验,由约定面达俗成的程度,为广大读者所普遍熟悉和承认。既然地图符号与概念之间存在的是约定的关系,那么就可以选择不同符号来指代某个概念。而不同符号之间又存在什么关系,则是地图符号的等价性问题。用地图符号代指概念,既然是一种约定,就可以用其他的图形去代替指示某一概

念。因此，地图符号的等价性指的是在地图符号代指概念的约定过程中，地图符号之间存在一种等价关系，即存在这样一个符号集合，集合中的每个符号都能代指同一概念。对地图符号的约定性和等价性的分析，可以将符号自身的特性和符号的实际应用，以及地图符号设计中的内部和外部作用规律区别开来。

②地图符号的分类。随着科学的进步和测绘事业的发展，地图符号的数量上升很快，过去的符号分类已显得片面和不完备了。例如，原先常把地图符号按视点的位置分为侧视符号和正视符号；按比例关系将地图符号分为依比例符号、不依比例符号和半依比例符号；按定位的情况将地图符号分为定位符号、说明符号和注记；按符号的外形特征将其分为几何符号、透视符号、象形符号和艺术符号等。

从地图符号的约定性及其表示的抽象概念，地图符号可分为点状符号、线状符号和面状符号。

——点状符号。当地图符号所指代的概念在抽象意义上可认为是定位于几何上的点时，称为点状符号。例如，测量控制点、居民点、独立地物、矿产地等符号。点状符号只能有一个定位点，而不管该点在符号的什么位置。对点状符号本身来说，该点是一数学上的点，但却可以代表实际呈点、线、面分布的事物。也就是说，不论地图符号是什么形状，什么尺寸，甚至形状有无意义，只要它只有一个有意义的定位点，则就是点状符号，因此各种个体符号是点状符号，各种统计图表也是点状符号。

——线状符号。当地图符号所指代的概念在抽象意义上可认为是定位于几何上的线时，称为线状符号。这时，符号沿着某个方向延伸，且宽度可以与地图比例尺没有关系，而长度与地图比例尺发生关系。例如，河流、沟渠、道路、等高线等符号。定位线可以对应于实际呈线状分布的要素，也可以没有实际的要素与之对应。例如动线符号有时可不对应实际的点位，它只表示某种方向，以及这种方向的变化过程。

——面状符号　当地图符号所指代的概念在抽象意义上可认为是定位于几何上的面时，称为面状符号。例如，湖泊、森林、植被、行政区域范围等符号。因此通常面状符号中的面与实际事物的面是对应的。地图表示方法中的范围法、质底法、量底法、分区统计图法都属于面状符号的应用。

③地图符号的构成。地图符号对表达地图内容具有重要作用，高质量的地图符号是丰富地图内容，增强地图的易读性和便于编绘地图的必要前提。从地图符号表示事物的质量和数量特点及视觉效果等方面来考虑，地图符号的构成可概括为图形、色彩和尺寸等三个基本要素。地图的清晰易读性在很大程度上是由这三个基本要素决定的。

· 符号的图形。符号的图形是决定符号的最根本、最主要的因素，它直接影响着地图内容的显示程度和效果，制约着各要素的相互配合及整个地图外貌的美观。GIS 中的地图制图对符号的要求是多种多样的，尤其在科学性和精度方面，提出了更严格的要求。此外，考虑到地图内容的完整性、地理适应性、艺术性及实用性等特征，又要求符号图形应简单形象化。地图符号图形的构成应从以下几个方面来考虑。

符号构图的图案化。从构图角度而言，要把实际地面上复杂对象的图形典型化，符号就不能是实地物体的真实写照，面必须用概括的方法，抓住物体的基本特征，用规则的图案表示出来。在这个过程中，要求舍去那些不必要的细微特征，做到尽量简洁，并用艺术手法使其美化，亦即"图案化"。

符号构图的逻辑性。在设计地图符号时，一定要与符号的含义建立起有机的、内在的联系，或者说让地图内容的分类、分级、重要、次要等变化也相应地反映为图上的变化，这样可以避免孤立地、片面地设计每个符号的图形，从而在不增加单个符号数量的前提下，丰富地图的符号系统。一般可以用图形的大小、粗细、实虚等来相应地反映要素的分类、分级、重要、次要等变化。

·符号的色彩。符号的色彩是构成地图符号的另一基本要素，它不仅可以丰富地图的内容，增大地图的载负量，使地图各要素层次分明、清晰易读，还能增加地图的艺术性。因此，色彩与图形融合在一起，有效地提高了地图的表现力。

在使用点、线、面符号的色彩时，面状符号常具有背景的意义，一般宜使用饱和度较小的色彩；点状符号和线状符号则常使用饱和度大的色彩，使其构成较强烈的刺激，而易为人们感知。在这个原则基础上，再结合色相、亮度和饱和度的变化，表现各种对象的质量特征、数量特征和分布范围等。

·符号的尺寸。符号的尺寸是构成地图符号的又一基本要素，它在很大程度上将影响地图各要素的显示程度及阅读能力。符号尺寸过大，会增加地图的载负量，减少地图的表示内容；过小则使读图者不易辨认，难以阅读，影响地图的清晰性。

地图符号的大小、线划的粗细、线划的间隔等是由多种因素决定的，最主要的因素是符号本身的形状、地图的主题、用途、比例尺及地图的载负量等，而且与绘图水平、印刷工艺有关。

一般情况下在不影响地图符号绘制时，都采用精细小巧的图案作为地图符号，以便在地图上表示更多的内容。此外，确定地图符号的尺寸还应考虑所表现事物和现象的主次，即重要的物体和现象用较大、较粗、较醒目的符号表示；而次要的、一般的物体和现象用较小、较细的符号表示。

④地图符号的设计。随着计算机技术和数据处理技术的发展，对于象纸质地图这样的传统形式模拟空间信息，已经越来越不适应用户对空间信息的要求，数字式空间信息已得到广泛应用。对数字式空间信息的各种分析和处理的结果，不能还以数字的形式呈现给用户，必须要转换为图形符号才能直观地表示。对于从数字空间信息转换成符号图形的过程(即符号化)，以往的做法是每套系统针对具体的任务开发一套符号化软件，一个符号用一段程序来实现。这种方法有很大的局限性，就是符号化软件与具体系统捆绑

在一起，当系统发生变化时，符号软件相应地也要变动，而且很难适应新符号的设计和制作。另外，用户在使用时，必须了解符号软件的结构及其提供的功能函数，了解这些函数的使用方法和要求，给用户带来很大的不便。因此，建立地图符号库系统就显得尤其重要。

地图符号库就是将常用的符号经分类整理后以数据库的形式存储到计算机中，实现其数据库的管理功能，用于符号信息的检索、存储、修改、定义和符号的重组。

地图符号库简单来说就是一个管理地图符号的数据库，它首先应该是一个结构完整、功能完备的数据库，同时还应该成为 GIS 的一部分(符号库管理子系统)。完整的符号库应保证在结构、功能方面具有一定的独立性，也就是说地图符号库不仅仅是一个存储地图符号的数据库，更应该具有符号建立、修改、删除、显示、查询等多种功能，这使得地图符号库可以独立于 GIS 以外进行研制和开发，更易于制作相对标准的符号库系统。但同时地图符号库也要便于纳入 GIS 中，成为其有效的组成部分，只有这样，地图符号库才具有使用意义，才能真正发挥它的作用。

因此，地图符号库的功能要包括以下几个方而：

- 各种地图符号的存储；
- 地图符号的建立；
- 地图符号的修改；
- 地图符号的删除；
- 地图符号的显示；
- 地图符号的查询、检索；
- 地图符号的输出、调用等。

其中，符号库中各种符号的建立、存储是其核心部分。

为了在地图上合适的位置输出一个地图符号，需要三类信息：

- 描述空间实体位置和形状的几何信息(定位点、定位线、边界线)。

·符号结构信息：描述符号本身的信息，包括符号由哪些图元（点、直线段、折线、圆等基本图形单元）组成和各图元的属性（颜色、大小、形状和方向），以及图元间的相互关系等信息。

·配置描述信息：确定如何依据几何信息在合适位置上输出大小、方向、形状和颜色符合要求的符号图形。

配置描述信息和符号结构信息一起共同确定了每个符号的特性，这两类信息统称为符号描述信息。地图符号库中存储的就是每个符号的符号描述信息，符号的区别仅仅在于它们的符号描述信息不同。

对于一个完整的地图符号库系统来说，除上述信息外，还必须有用于理解符号描述信息的解释模块和用于形成符号图形的图形生成模块。如果用 x 和 y 表示几何信息和符号描述信息，用函数 f 表示信息解释，用函数 S 表示图形生成，则符号化过程的函数表示为

$$G=S(x,f(y))$$

其中：G 为符号化的结果，即符号图形。用以往的做法对符号直接编程，将上式用一个函数表示就变为

$$G=F(x)$$

显然，符号系统的任何改变就意味着 F 的变化，即引起程序的变化，所以要建立不涉及程序变动就能改变符号系统的生成软件，就必须把符号描述信息 y 从程序中独立出来，和 x 一起作为程序的输入参数。函数表达式为

$$G=P(x,y)$$

其中 P(x,y)为包括信息解释和图形生成功能的信息处理过程。

在上式中，x 是来自地理数据库的空间目标的几何坐标信息，函数 P(x,y)能够通过编程实现。对于空间地理信息符号化过程，能否成功实现的关键是 y（即符号描述信息）是否可定义，因为符号系统的定义和修改完全体现在 y 的定义与修改上。我们通过对

点、线、面符号构造原理的分析，证明它是可以定义的。

目前，地图符号库的设计正是基于以上原理，将描述符号的参数和结构信息与解释这些参数并输出符号的程序相分离，符号的建立、更新与增加，不应通过重新编程来实现，而是通过简单的人机交互操作实现。

§9.2 地图符号的设计与管理

地图按主题(内容)可以分为两类，即：普通地图和专题地图。普通地图是以相对平衡的详细程度表示水系、地貌和土质、植被、居民地、交通网、境界、管线等基本地理要素的地图。它比较全面地反映制图区域的自然环境、地区条件和人类活动的一般状况，也反映出自然、经济诸方面的相互联系和影响的基本规律。专题地图是根据专业方面的需要以一种或几种要素为主题的地图，作为专题地图的要素，可以是普通地图上有的，也可以是普通地图上没有但属于专业部门特殊需要的内容。

由于普通地图和专题地图在表示内容上有很大的不同，所以这两种地图所采用的表示方法也不同。普通地图是运用普通地图符号系统，根据一定的数学法则，绘制在平面上的图形。专题地图上为了表示专题要素的现象和反映这些要素的实际指标，必须运用专题地图符号系统，采用定点符号法、线状符号法、定质区域法、等值线法、点值法、动线符号法、等值区域法和统计图表法来表示各专题内容。

1. 普通地图符号的设计与管理

(1)点状地图符号的设计与管理。

①点状符号的特点和构图规律。点状符号是用来表示其定位点上的地图信息的地图符号，如水塔、烟囱、测量控制点等符号。点状符号除了拥有地图符号的一般特点外，还有许多自己的特点：

- 点状符号图形形体相对较小；

· 点状符号图形都有准确的定位点和方向性；

· 点状符号图形固定，不随它在图幅中位置的变化而变化；

· 点状符号图形大小不随比例尺的变化而变化；

· 点状符号图形比较规范，大都是由几何图形构成。

在地图上，通常以形状变量来表示实地物体或现象的质量差别，以尺寸变量来表示物体或现象的数量差别。形状变量是构造点状符号的主要符号变量，所以在这里只讨论点状符号在形状方面的构图规律。根据点状符号图形形状的不同，可粗略地把点状符号分为两大类：象形符号和几何符号。

象形符号是以简化的形式模拟制图物体的外观，具有形象直观的特点。采用象形符号来表示客观物体，可以使人"望图思物"，无须看图例就知道该符号所代表的物体，因此象形符号深受大家的欢迎。但是由于象形符号有时过于刻意模仿客观物体，导致其图形形状较为复杂，计算机实现起来比较困难，而且还不容易精确定位，故在地图中尽量少用象形符号，且其图形也要越来越简化。

几何符号是用简单的几何图形及其组合来指代客观物体的符号，它具有简单易绘、容易定位的特点。几何符号在地图上的种类很多，其在构图方面的变化也多，很容易对不同的符号进行区分。几何符号的构图规律有以下几条：

——利用基本形状作为符号主体，容易区分。地图上的许多点状符号都是利用三角形、正方形、圆形等基本几何形状作为符号的主体部分，使得符号的形状有较大差别。

——使用结构元素，赋给符号语义以种属和群的概念。地图上有些点状符号意义相近，或有着某些特殊的联系，则使这些符号在结构上有一个共同的结构元素，以突出符号之间的联系。如"土堆上的三角点"与"三角点"有共同的结构元素——三角形。

——利用符号与注记相结合的方法来表示某种具体事物。有些点状符号是一个符号可以表示多种物体，只有通过符号旁边的注记才能区分具体是哪一种。

②点状符号的构造原理。点状符号可看成是一个有限直线段的集合，各直线段通过统一的坐标系联系在一起。但是，有些点状符号如果只通过直线段的集合来描述，将是很困难的。例如，圆用直线段进行人工拟合来表示就比较困难，而用圆的参数定义就比较容易，只需圆的半径、圆心的定位点和圆的色彩。为了方便点状符号的设计，定义点状符号为任意线段和规则几何图形的组合。因此，一个点状符号的设计可先定义一个坐标系，坐标系的原点就是此点状符号的定位点。如图9—3所示，此符号可定义为集合：{圆，圆心(0,0)，半径(1)，色彩，宽度}U{直线段，起点(－1,2)，终点(1,2)，色彩，宽度}。此处的1代表一个标准单位。

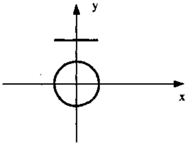


图9—3 符号坐标系

任意线段和规则几何图形称为图元(Cell)。一个点状符号就以图元为基础进行设计，设计时以一个统一的坐标系(符号空间)为准，坐标系的原点就是此点状符号的定位点。符号定位点的定义并不统一，但在标准图式中都有明确的规定，定位点的选取必须遵从图式规定。点状符号的符号描述信息，即点状符号中图元的大小、位置、相互关系，是由图元控制点与符号定位点间的关系推导出来的。

图元应当是点状符号中常见的规则几何图形。所以在设计点状符号之前要先定义一些基本的图元，一个复杂的点状符号就是由基本图元构成。点、直线段、折线、圆、椭圆、圆弧等都是点状符号中常见的规则几何图形，可以把它们定义为基本的图元。通过对地

图图式中点状符号结构的分析，符号制作系统中用来构造符号的基本图元可以是以下九种：

点、直线、折线(或多边形)、曲线、圆、椭圆、弓形、扇形、文字。

利用以上图元，进行合理的组合，基本上可以构造出地图图式中的所有点状符号。如图 9—4 所示，水塔符号由四个直线段图元和一个折线(封闭)图元构成，油库符号由一个半圆弧(特殊弓形)图元和一个填充半圆(特殊扇形)图元构成，棕榈树符号由三个直线段图元和一个曲线图元构成。



水塔　　油库　　棕榈树

图 9—4　点状符号的图元

但为了设计出理想的点状符号，还应遵守一定的原则：

——明确各图元的定义参数。前面介绍的九种图元是点状符号设计的基础，它们的定义参数各不相同，要用最节省存储空间的参数来定义各图元，同时也要有利于图元的绘制。在下面对图元数据结构的设计中还要详细论述。

——选择最少数量的图元。在对具体的点状符号进行设计制作时，要尽量选择最少数量的图元，以利于点状符号的设计和点状符号库的管理。

——点状符号设计时应有统一的坐标系。各个点状符号的定位点坐标都是以这个统一坐标系为标准。

③点状符号库的数据结构设计。点状符号库包含了两个部分：点状符号的描述信息参数集和点状符号中图元的绘制程序。点状符号参数集的结构，就是点状符号库的数据存储结构。

点状符号的符号描述信息是可以定义的。显然，构成点状符号

的图元个数对于某一个特定符号来说是个实数。符号的描述信息则表现为各图元与符号定位点之间的关系,它通过图元的各控制点(如多边形的角点、圆的外接矩形和线段的端点)与符号定位点的关系反映出来。对于方向可变的点状符号(如桥梁),符号描述信息还应包含各图元与符号方向线(符号定位点与方向点的联线)的关系,这一关系也可通过图元控制点与符号方向线间的关系间接表示出来。

用两个文件存储点状符号库:点状符号库索引文件和点状符号库数据文件其结构举例说明如下:

——点状符号库索引文件

符号代码	符号地址	符号存储长度	符号定位点	符号删除标志
int(4)	int(4)	int(4)	CPoint(8)	int(4)

每个点状符号在索引文件中占 24 个字节的存储空间。其中符号代码是符号的标识,用来区别各符号。符号地址是该符号在数据文件中存储的起始位置(以字节计),符号存储长度是该符号在数据文件中所占的字节数。符号定位点数据项用一个 CPoint 数据结构存储该符号的定位点 x,y 坐标。符号删除标志用来对一个符号在逻辑上是否删除进行标识,但是在逻辑上已经删除的符号在物理上并没有删掉,即它的索引和数据仍然在索引文件和数据文件中,必要时还可以对删除的符号进行恢复。要想进行物理的删除,可以在符号库的整理时,把已经作删除标识的符号的索引和数据从相应的文件中删掉。

——点状符号库数据文件

图元码 1	参数 1	参数 2	………	参数 n_1
图元码 2	参数 1	参数 2	………	参数 n_2
图元码 3	参数 1	参数 2	………	参数 n_3

其中图元码用来对各图元进行标识,例如点图元用 0 来标识、

直线图元用 1 来标识。对于不同的图元来说，它们的参数个数和参数的类型都各不相同。每个符号在数据文件中占有的存储空间是不相同的，取决于符号的图元个数、图元类型和图元参数的长度。

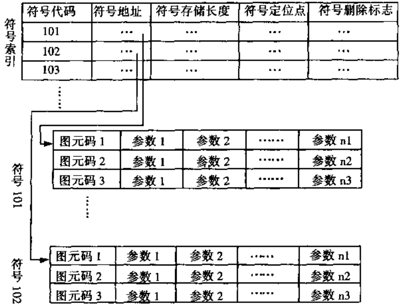


图 9—5 符号的索引与数据的联系

符号库中索引文件与数据文件的联系如图 9—5 所示：

在绘制点状符号的时候，只需根据点状符号库索引文件中的符号代码、符号地址、符号存储长度和定位点坐标等信息，到点符库数据文件中取得相应符号的数据（符号描述信息），再将该符号的尺寸变换为所需要的尺寸后，把各点对于符号定位点的坐标加上该符号对应的点状地理目标的实际坐标（x_0，y_0），配置到地图空间即可。

（2）线状地图符号的设计与管理。

①线状符号的特点和构图规律。线状符号是用来表示地图上长度依比例，而宽度不依比例的顺线状延伸分布的地物，如河流、

沟渠、道路等符号。线状符号除了拥有地图符号的一般特点外，还有一些自己的特点：

· 线状符号都有一条有形或无形的定位线；

· 线状符号可进一步划分为曲线、直线、虚线、点状符号线等；

· 线状符号的图形可看作若干图形组合而成，例如虚线是由短直线和空白段组合而成。

构成线状符号的符号变量一般是形状、色彩、尺寸和图案。形状和色彩变量用来表示物体的质量差别，尺寸变量用来表示物体的数量差别。由于线状符号在地图上的面积不可能很大，用图案来表示效果并不好，所以，线状符号很少使用图案符号变量。

线状符号的形状变量指的不是线状符号的整体形状，而是指构成线状符号的基本单元的形状。这个基本单元是一个区段，它可以由一些简单几何图形（图元）组成，线状符号就是多个同样形状的区段沿定位线循环配置而成的。因此，线状符号形状的差别就体现在线状符号循环段的形状区别上。线状符号的尺寸变量包括线状符号的宽度和线状符号的长度，线符宽度可以定义为循环段宽度的倍数，线符长度就是符号循环段的长度。线状符号的色彩变量是整个线状符号的颜色，而不是循环段中图元的颜色。这样，形状相同而只是尺寸或颜色有区别的线状符号被认为是同一个符号，在符号库中也只存储一次，只是在绘图调用时改变符号的尺寸或色彩。因此，在线状符号库中只保存符号的形状，对于符号的尺寸变量和色彩变量，只是作为外部调用的参数，这样既可以节省存储空间又方便用户的使用。

②线状符号的构造原理。为了确定线状符号的描述信息，可以把线状符号分为四种情况，如图 9—6 所示。

这样对于构成某一线状符号的图元个数及图元本身就能确定下来。线状符号的图元与定位线的关系，如平行、重合、垂直、有夹角、定位线的起点、终点等，是通过记录图元的位置（位于起始端、终止端或循环区域）和图元的控制点与定位线直控线段的坐标关

系来反映的。定位线直控线段定义为与符号定位线重合的线段，它的长度与线状符号循环段的长度相等，对于图 9—6 中(1)的情况，其长度定义为“定位线之长”；而图 9—6 中(2)的情况，其循环段的长度定义为从一个实短线开始到下一个实短线起始处的距离。对于线状符号的一个循环段来说，它其实就是一个点状符号，只是该点状符号的定位点在每个循环段的起点，方向与线状符号定位线直控线段重合。因此，构成线状符号的图元与点状符号类似，也是点、直线、折线(或多边形)、曲线、圆、椭圆、弓形、扇形、文字等九种。

在绘制线状符号的时候，首先要经过“定长提取”获得每个循环段的起、终点和拐弯点坐标，然后以它的起点为绘制符号的定位点，以它的定位线直控线段的方向为循环段内各图元的旋转方向，绘制该符号的各个图元。若某个图元超出了前方拐弯点，则截去超出部分，将截去部分转到下一折线段内处理。对有截去情况的图元，在拐弯点处还要做变形处理，使得图元随定位线弯曲，在拐弯点处能够紧密结合，而不出现裂缝或重叠现象。

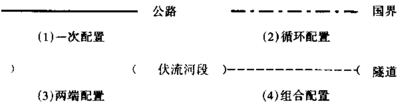


图 9—6 线状符号分类

构成线状符号的图元有九种之多，这些图元中有的图元在线状符号的循环段中所占的长度很小，例如圆、扇形、文字等图元。对于这样的图元，当位于循环段的拐弯点时，只能改变它的方向而不能改变形状。对于那些面积很小的图元，也可以在拐弯点处将它直接绘制出来，或只是做一下移位处理，将它的位置从拐弯点处移开再绘制。而对于像直线、折线、曲线等图元，在线状符号的循环段中

所占的长度可能很大，所以在绘制的时候要将超出循环段拐弯点的部分截去，将截去部分转到下一折线段内处理，并在拐弯点处做变形处理。包含这些图元的线状符号绘制过程的算法参见图 9—7、图 9—8。

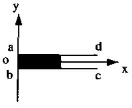


图 9—7 线符的循环段

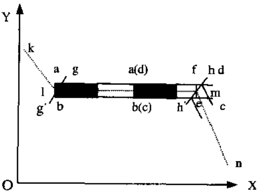


图 9—8 线符的绘制过程

地图线状目标的定位线为 k—l—m—n，在图 9—8 中用虚线表示，地图坐标系为 XOY。线符循环段所在的局部坐标系 xoy(局部坐标系原点 o 与节点 l 重合)，循环段所有图元的外接矩形为 abcd，节点 l 和 m 处的角平分线分别为 gg' 和 hh'，循环段在节点 m 处被分割为两段：abef 和 fecd。分割线 ef 通过 m 且垂直于 lm。两条角平分线和分割线将循环段分割为三个部分，并作不同处理：

—— 在局部坐标系 xoy 中，若循环段中图元的控制点 p(xp，yp)满足 0<xp<xm，且点 p 在两角平分线之间，则该点不作变形处理，只按一般方法从局部坐标系变换至地图坐标系。

——在局部坐标系 xoy 中，若循环段中图元的控制点 p(xp，yp)满足(xp =0)或(xp =xm)或(0<xp<xm 且 p 点不在两角平分线之间)，则该点沿 x 轴正向或反向平移至附近的角平分线上，然后变换至地图坐标系。图 9—8 中三角形 alg，fmh 内的点便属于这种情况。

——图元超出前方节点 m 的部分，即四边形 fecd 内的点 p（xp，yp），在局部坐标系中满足 xp≥xm，则将这些点转入下一节 m－n 中进行处理，重复以上过程，直至符号全部完成。

③线状符号库的数据结构设计。由上面的分析得出，线状符号的描述信息是可以定义的，而且对于每个循环段来说，与点状符号的符号描述信息类似。所以，线状符号库的存储结构与点状符号库类似。

同样用两个文件存储线状符号库，即线状符号库索引文件和线状符号库数据文件，这两个文件的数据结构如下所示：

——线状符号库索引文件

符号代码	配置方式	符号地址	符号存储长度	删除标志
int(4)	int(4)	int(4)	int(4)	int(4)

每个符号在索引文件中占 20 个字节的存储空间。其中符号代码是符号的标识，用来区别各符号。配置方式是区分符号类型的标识，0 代表一次配置，1 代表两端配置，2 代表循环配置，3 代表组合配置。符号地址是该符号在数据文件中存储的起始位置（以字节计），符号存储长度是该符号在数据文件中所占的字节数。符号删除标志用来对一个符号在逻辑上是否删除进行标识。

——线状符号库数据文件

图元码 1	位置码 1	参数 1	参数 2	……	参数 n_1
图元码 2	位置码 2	参数 1	参数 2	……	参数 n_2
图元码 3	位置码 3	参数 1	参数 2	……	参数 n_3

每个符号在数据文件中占有的存储空间是不相同的，取决于符号的图元个数、图元类型和图元参数的长度。其中位置码表示该图元的位置特征，位置码为 0——图元位于循环段，位置码为 1——图元位于线要素的首端，位置码为 2——图元位于线要素的末端。

(3)面状地图符号的设计与管理。

①面状符号的特点和构图规律。面状符号是指地图上用来表示呈面状分布的物体或地理现象的符号，它在二维图的各方向上都能依比例表示，如依比例表示的耕地、果园、水域等符号。面状符号有如下特点：

• 一般有一个有形或无形的封闭轮廓线；

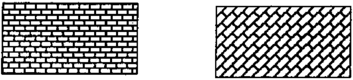
• 为区别轮廓范围内的对象，多数面状符号要在轮廓范围内配置不同的点状、线状符号或普染颜色。

面状符号由面状符号的面和面状符号的轮廓构成，面状符号的面由符号变量的色彩(指底色)和图案构成。色彩通过不同的色相表示要素的质量差别，而用明度和饱和度的变化来反映要素的数量差别。图案也可以表示面状符号的质量和数量差别，图案本身也有色彩，这是指图案的色彩，而不是指底色。有时面状符号可以用色彩(底色)加图案来表示，这主要是用于强调，例如在用不同的图案分类后，又给每种图案以不同的色彩以加强分类的效果。

面状符号的色彩在计算机上通常用 RGB(红、绿、蓝)来表示，但用户很难用该颜色空间描述心理颜色，面是用 HLS(色相、亮度、饱和度)来对色彩进行设计。对于面状符号的图案，最常用的方法是在多边形内以井字形或品字形配置点状符号和不同类型的晕线。

②面状符号的构造原理。为了确定面状符号的描述信息，可以把面状符号分为两类，如图 9—9 所示。

第一类面状符号称为位图填充面状符号，它是由一个 8×8 的位图模板来定义各个面符，再由这个模板按行、列顺序排列，填充整个多边形区域面形成的。位图中的每一个位点对应屏幕上的一个像素，位图的设计制作可以在一个 8×8 的网格上进行，用户可利用面符编辑器中的工具设计制作任意形状的位图。一个位图模板由一个长度为 8 的整数数组来定义，数组中的一个整数对应位图的一行。



(1)位图填充面状符号



(2)点符填充面状符号

图 9—9　面状符号分类

第二类面状符号称为点符填充面状符号，它是在多边形内以不同的填充方式（井字形、品字形、散列式等）配置点状符号而形成的。实现的方法是：先根据配置点符的间隔距离，求出不同填充方式下配置点符的定位点坐标，然后在定位点上绘制该点符。点状符号的描述信息来自点状符号库，在面状符号中只需包含一个点状符号的符号代码即可。

③面状符号库的数据结构设计。用两个文件存储面状符号库，即面状符号库索引文件和面状符号库数据文件，这两个文件的数据结构如下所示：

——面状符号库索引文件

一个面状符号库只能关联一个点状符号库。

关联的点符库名称
CString

——面状符号库数据文件

符号代码	int(4)
填充方式	int(4)(0:点符填充,1:位图填充)

点符代码	int(4)
排列方式	int(4)(0：井字型,1：品字型,2：散列式)
位图数组	short * 8(16)
删除标志	int(4)

其中符号代码是符号的标识,用来区别各符号。填充方式是区分符号类型的标识,0代表点符填充,1代表位图填充。点符代码是用来标识进行点符填充的点状符号。排列方式用来标识点符填充中配置点符的方式,0代表井字形排列,1代表品字形排列,2代表散列式排列。位图数组是用来存放位图填充的位图模板的定义数组。删除标志用来对一个符号在逻辑上是否删除进行标识。每个面状符号的物理存储长度是固定的,在面状符号库的数据文件中一个面符占36个字节。

2. 专题地图符号的设计与管理

在GIS的专题地图制图系统中,可以给出成百上千种的地图符号供用户选择,但大多数用户缺乏或不熟悉地图表示方法的有关知识,因面很难做出正确的选择。所以,GIS中的专题地图制图系统不能只是符号的绘制,而应包括地图符号的正确使用。例如,可以根据地图数据的特点,自动提供或建议用户使用某一种地图符号,这样可以为不熟悉地图设计原理的用户提供极大的方便。因而对于专题地图符号设计与管理讨论的重点,不是某一种地图符号的具体绘制方法,而是介绍地图符号在绘制、使用、管理等方面的特点和要求。

(1)点状专题符号的设计与管理。

①点状符号所表示的地图要素。

——单要素和联合要素。所谓联合要素是指由几个具有一定关系的定量要素(单要素)组成的地图要素,并且可以用一个地图符号表示。例如,当表示工业总产值和农业总产值时,可以作为一个要素,即工农业总产值。联合要素又分为结构型和非结构型。

结构型联合要素包含的子要素反映了一种结构，如工农业总产值反映了工业总产值和农业总产值的总和以及各自所占的比例；而非结构型联合要素的子要素不表示结构，而只表示一种关系，如粮食总产和粮食亩产。

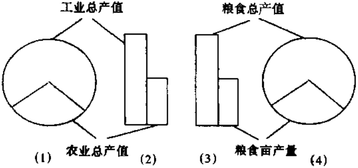


图 9—10 结构型和非结构型联合要素

之所以要把联合要素分为结构型和非结构型，是因为在符号设计时要考虑它们之间的差别。结构型联合要素通常应以结构性的符号表示，如工农业总产值可表示为图 9—10(1)，而用非结构性符号(如 9—10(2))表示时，则效果不好，因为没有表示出要素的结构性。反之，非结构型联合要素应以非结构型符号表示(9—10(3))，而用结构性符号表示(如 9—10(4))时，会使读者产生错误的理解。

所谓单要素是相对于联合要素而言的，即指一般的地图要素。联合要素可以看作是包含了几个单要素的地图要素，而且要用一个符号表示，所以联合要素和单要素在符号设计上的要求是不一样的。

——点状符号所表示的地图要素的类型。点状符号是只有一个有意义的定位点的地图符号。点状符号不仅可以表示点状要素，而且还可以表示顺序分级、间隔分级和数值表示的面状要素。为了使点状符号与地图要素有明确的关系，并考虑到点状符号在设计上的不同，可以把点状符号所表示的要素分为五类：

质量型单要素：单要素的定性表示、分类表示和顺序分级。

数量型单要素：单要素的间隔分级和数值表示。

时间型单要素：单要素表示趋势时的间隔分级和数值表示。

结构型联合要素：子要素具有结构性的定量要素（间隔分级和数值表示）。

非结构型联合要素：子要素不具有结构性的定量要素（间隔分级和数值表示）。

以上五种类型概括了绝大多数可以用点状符号表示的地图要素。

②质量型单要素的点状符号的设计与管理。质量型单要素所用的点状符号如图 9—11 所示：

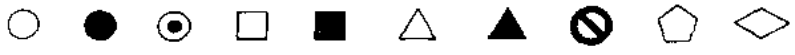


图 9－11　质量型单要素符号

这些符号通常称为个体符号。这些符号通常用形状反映要素的定性和分类，而用尺寸反映要素的顺序分级。由于点状符号尺寸较小，所以色彩通常用来强调定性和表示分类，用于表示等级效果不好。单独用形状来表示顺序分级的效果也不好，通常可用形状结合尺寸的方法。

这种类型的点状符号在计算机地图制图系统中有两种形式，即矢量型和栅格型。栅格型点状符号记录为点阵信息（类似于点阵汉字的存取方式），因而很容易建立和管理，但用矢量绘图机输出时则很麻烦，而且在 GIS 中进行放大、缩小时的效果不好，因此常用矢量型点状符号。

在 GIS 的专题地图制图中，表示质量型单要素的点状符号的设计与普通地图的点状符号类似，即建立点状专题符号库。表示质量型单要素的点状符号都是由规则的几何图形（三角形、圆形、折线、点等）构成的，这些基本的几何图形称之为图元，每个图元都具有形状、色彩、尺寸三个视觉变量。因此，点状符号可以通过描绘图

元之间的关系，和说明点状符号的定位点来完全确定。

具体的点状符号设计大多通过人机交互手段，由专门的软件实现。通常是在屏幕上显示一矩形格网以表明符号的设计区域，而在屏幕的另一部分显示图元，用户可以用鼠标或键盘来选择构成符号的图元，并可以改变其尺寸和色彩，移动图元使之处在符号设计区的合适位置后认可，再取点状符号的下一个图元作同样的处理，直到一个符号设计完，便把符号的描述信息记录下来。这类软件可生成复杂的点状符号，并且可以方便地进行修改。

这种方法设计出的符号在形状确定的同时，色彩被固定了，尺寸只能用符号整体的放大或缩小来调整。这样的话，对于形状和尺寸相同，而色彩不同的符号则需设计成不同的点状符号，这对于用色彩分类的点状符号和用色彩辅助尺寸分级的点状符号的使用将产生相当的困难。因此，总的来讲，这种符号设计方法更适合于定性要素的表示，对于用符号的形状分类的情况和单纯用尺寸分级的情况也可适用。为了适应形状不变而色彩改变的情况，可对这种方法作某些修改，如把符号设计时的色彩作为缺省色彩，而允许在使用符号时给出一种色彩以取代原来符号的某个或某些图元的色彩，这样就可方便地使用形状、色彩和尺寸三个视觉变量来表示单要素的定性、分类和顺序分级了。

对表示质量型单要素的点状符号，在使用时应通过定位点坐标、符号的形状、色彩和尺寸来确定一个具体的点状符号。在这里，符号的形状可以通过形状的编号确定，色彩可以用红、绿、兰所占的比例确定，而点状符号的尺寸在反映要素的等级时，就是指符号的大小，因而只应该有一个尺寸，当符号的尺寸改变时，其图元的尺寸应自动作相应的变化。

③定量型要素的点状符号的设计与管理。定量型要素是指间隔分级和数值表示的地图要素，包括数量型单要素、时间型单要素、结构型联合要素和非结构型联合要素，相应的点状符号由图9--12 所示。

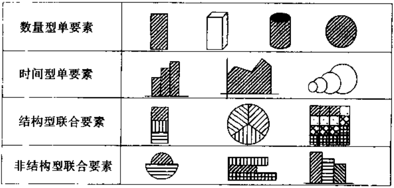


图 9—12　定量型要素的符号

用计算机对这些点状符号进行设计和管理时，通常是以编程的方法来实现的。

对于表示数量型单要素的点状符号，其形状往往与它所表示的要素有关，例如习惯用圆来表示区域的人口数。对这种点状符号，最重要的视觉变量是尺寸，可以用以反映要素的数量。当表示要素的间隔分级时，符号的尺寸表示了要素的等级，因此尺寸之间需要有一定的关系；当表示要素的精确数值时，符号的尺寸将随数据的不同而改变。因此，对这种表示数量型单要素的点状符号，在给出了定位点、形状编号、色彩或图案和一个符号的尺寸后即可绘出。

时间型单要素是指随时间变化的地图要素，如人口逐年的变化。对于表示时间型单要素的点状符号，其符号的形状是指构成符号的形式，所以图 9—13 所示的符号实际是同一个形状。也就是说，表示时间型单要素的点状符号可以看作是多个表示数量型单要素的点状符号的合并，其数目在该种符号的绘制程序中应该是个变量，在具体使用该点状符号时与地图要素有多少不同时间的数据相对应。色彩或图案通常用来反映地图要素的性质，而这种点状符号表示的是单要素的变化趋势，因此色彩或图案只有一种。而

该种点状符号的尺寸数目也应该由地图要素的数据的多少来确定，而具体的尺寸将根据地图要素的具体数据来确定。例如，要表示我国各省、自治区、直辖市近五年人口数的变化时，则每个省将有五年的人口数需进行表示，可以使用图 9—13 所示符号，即指定使用这种形状的点状符号，同时，还需给定色彩，例如用红色，而最重要的则是给出该符号的五个尺寸。

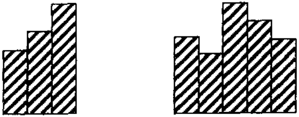


图 9—13　时间型单要素符号

当用间隔分级表示人口数时，需首先把人口数进行分级，并确定每级对应的符号的尺寸，然后根据某地区五年的数据分别确定属于哪个等级，这样就可以确定这个符号的各部分的尺寸。在给出了定位点后，就可绘制符号了。

结构型联合要素是指联合要素的子要素反映了一种结构，而子要素的多少和点状符号的形状没有关系。子要素之间实际是分类的关系，所以通常用色彩来表示子要素的分类，有时也可用图案来对子要素进行分类，但在彩色地图上较少使用。对这种表示结构型联合要素的点状符号，其尺寸是对整个符号而言的，因而间隔分级时符号的尺寸就是指整个符号各个等级的尺寸，而数值表示时，具体某个符号的尺寸可由它所表示的数据与最大尺寸所表示的数据的比例来确定。在确定了整个符号的尺寸后，还要确定符号内部各子要素所占的比例，这只要根据具体的数据就可计算出。因此对绘制这种点状符号的软件，需要提供符号形状、子要素个数 n、子要素对应的色彩或图案 n 个、符号的总体尺寸、子要素各占的比例

和定位点的坐标才能绘制出相应的符号。



图 9--14 非结构型联合要素符号

表示非结构型联合要素的点状符号与表示结构型联合要素的点状符号在设计时有许多相似之处,如需要用色彩表示子要素之间的分类关系等,主要区别在于:对非结构型点状符号,尺寸是对各子要素而言的,因此对间隔分级的非结构型联合要素要对每个子要素进行分级,而且要给出每个子要素各个等级相应的符号尺寸。例如对表示粮食总产和粮食亩产的联合要素,假设需要对粮食总产分五级,对粮食亩产分六级,并用图 9—14 所示符号表示,则需对亩产的六个等级分别确定尺寸,对总产的五个等级也分别给出尺寸。而某一个符号只需要两个尺寸,分别由这个符号所表示的亩产和总产的等级确定。所以,当给定了符号的形状、子要素的个数 n,相应的色彩或图案 n 个、尺寸 n 个和定位点的坐标后,即能绘制非结构型的点状符号。

通过上述分析可以看出,虽然点状符号是由形状、色彩、尺寸和图案视觉变量构成,但根据要素的特点,其相应符号的视觉变量的含义有所不同,例如对于尺寸,有时是指符号的总体尺寸,有时是指符号各部分的尺寸,但只要弄清了它所表示的地图要素的类型,就能知道每个视觉变量的含义,就能进行具体的符号设计和绘制了。

(2)线状专题符号的设计与管理。

①线状符号的类型及其视觉变量。构成线状符号的视觉变量是形状、色彩、尺寸和图案,线状符号的设计就是确定构成线状符号的视觉变量的具体值。



(1)　　　　　　　　(2)

图 9—15　线状符号的尺寸变化

视觉变量原理认为尺寸可以表示顺序和等级，但对线状符号来说，“尺寸”的具体含义是什么呢?图 9—15 中的(1)图是虚线。虚线的加宽是尺寸的变化，每段线的长度变化是尺寸的变化，虚线的间距也是尺寸的变化，因此虚线有三个尺寸的变化。图 9—15 的(2)图是一种境界符号，具有更多的尺寸变化。但是我们可以发现，并不是每个尺寸的变化都意味着顺序和等级的变化，所以，可以认为线状符号的尺寸可以分宽度尺寸和长度尺寸两部分。宽度尺寸是控制线状符号宽度方向各部分的尺寸，长度尺寸是控制线状符号长度方向上各部分的尺寸，而且这两个尺寸的变化都可以表示顺序和等级的变化，因此可以认为这两个尺寸是线状符号的视觉变量“尺寸”的具体体现。因此当图 9—15 中(2)图中的宽度尺寸变化时，其短线的长度、线段的宽度、圆点的半径都要作相应的变化。

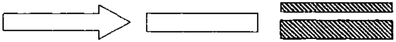
用于表示定性、分类和顺序分级的线状符号我们称为质量型线状符号，用于表示间隔分级和数值表示的线状符号称为数量型线状符号。质量型线状符号由形状、色彩和尺寸三个视觉变量构成，数量型线状符号由形状、色彩、尺寸和图案四个视觉变量构成，且两种线状符号在计算机地图制图中的绘制和处理是不一样的，所以分别介绍如下。

②质量型线状符号的设计与管理。这种线状符号在设计时与普通地图的线状符号类似，可以建立线状专题符号库。利用符号库在绘制线状符号时，如果把符号每部分的尺寸都定死，那就成了固定的线状符号，显然不行。但如果把线状符号每部分的尺寸都作为调用参数，则在使用时就很不方便。一个较好的办法是把线状符号的各部分的尺寸都转换成随宽度和长度尺寸变化的变量，这样的话，在调用绘图函数绘制线状符号时，只要给出定位线坐标、符号

编号、色彩和尺寸(宽度和长度尺寸)即可,大大方便了使用。

同点状符号一样,也可以通过类似的人机交互手段来定义和设计线状符号,而且修改和增删都很方便。为了使通过线状符号的描述信息来定义的线状符号在使用时更加方便和灵活,调用某一线状符号时,应允许以参数的形式修改它的色彩和尺寸(宽度尺寸和长度尺寸),这样才能使线状符号的使用简单而方便。

③数量型线状符号的设计。表示间隔分级和数值表示的地图要素的线状符号有多种,图 9—16 列出了常用的三种。当需要表示方向时用(1),无方向时用(2),双向时用(3)。例如要表示石油的流向时,常使用(1)所示的线状符号,而且往往用较暗的色彩表示"石油"的运输量。再如要表示道路双向的车流量时,则可用(3)所示的线状符号。



(1)　　(2)　　(3)

图 9—16　数量型线状符号

数量型线状符号由形状、色彩、尺寸和图案四个视觉变量构成,常用尺寸来表示等级和具体的数值。

数量型线状专题符号的绘制往往用软件编程实现,一种符号对应一段程序。

(3)面状专题符号的设计与管理。

①面状符号设计的基本要求。面状符号由面状符号的面和面状符号的轮廓构成,面状符号的面由视觉变量色彩(指底色)和图案构成。色彩通过不同的色相表示要素的定性和分类,而用明度和饱和度的变化来反映要素的顺序和等级,有时也使用色彩的其它性质,如用色彩的远近感来表示要素的顺序和等级。图案也可以表示面状符号的定性、分类和分级,图案本身也有色彩,这是指图案的色彩,而不是指底色。有时面状符号可以用色彩(底色)加图案来

表示，这主要是用于强调，例如在用不同的图案分类后，又给每种图案以不同的色彩以加强分类的效果。

对于表示定性的要素，只要根据指定的色彩或图案对要素相应的坐标数据（即多边形）进行填充即可；对于分类，首先要确定用哪些色彩或图案来进行分类，再根据某一要素具体属于哪一类，选择相应的色彩或图案进行填充；如果表示分级，则首先要确定用哪些色彩或图案来构成等级，然后根据每个要素的等级选用相应的色彩或图案进行填充。对于政区图的填色或类似的情况，理论上只要有四种色彩就足够了（如果没有飞地的话），在实际制作政区图时，通常用五种以上的色彩。对政区的设色要求每两个相邻的政区都用不同的色彩区分开，如果用计算机自动设色，则需要进行逐步的尝试才能完成，计算量很大，而且很难指定某一政区用某种色彩，对于象广东省和海南省这样，虽然没有相邻的陆地边界，也要着不同色彩等特殊情况很难处理，因此大多数采用人机交互的手段。

面状符号的轮廓可以看作是线状符号，由形状、色彩和尺寸视觉变量构成，因而可以当作线状符号来处理。

②色彩的设计。计算机上色彩通常用 RGB（红、绿、蓝）表示。RGB 空间是一个三维正立方体（如图 9－17）。

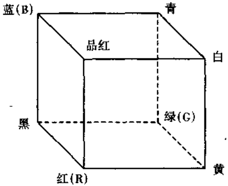


图 9－17 色彩的 RGB 空间

它的三个基本轴分别表示 R、G、B，原点为黑色，白色位于黑色的对角顶点。用户很难用该颜色空间描述心理颜色。因为人类大都具有给颜色指定色相、亮度和饱和度的能力。当给出颜料让人调配出指定的颜色时，他一定会先调 配出类似的色相，再加白色或黑色来调整其亮度和饱和度。因此色相、亮度和饱和度构成的颜色空间（即 HLS 空间）对色彩的设计更有用。HLS 空间是个双圆锥体（如图 9—18 所示）。RGB 空间可以方便地转换到 HLS 空间中。

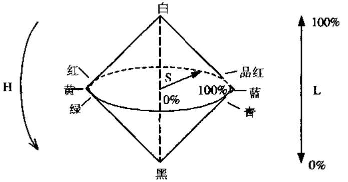


图 9—18　色彩的 HLS 空间

有了 RGB 和 HLS 空间的相互转换后。就可在计算机上用色相、亮度和饱和度来定义一个颜色了，这样更符合人的视觉感受。所以，地图屏幕色的设计可以通过人机交互的手段，通过分别修改色相、亮度和饱和度来得到需要的色彩，而在计算机内部则需把色相、亮度和饱和度转换成 R、G、B 值。

§9.3　地图注记与地图图例配置

1. 地图注记配置

地图注记是一种广义的地图符号，用以配合其他符号说明事物的名称、数量、质量特征，是地图中不可缺少的重要内容。

(1)地图注记配置的研究现状。

①注记的重要性。地图注记是地图的重要组成内容。借助于地图注记，读图者不仅可以识别图上的重要目标，而且还能获取目标的其他相关信息。不难想象，如果一幅地图上没有任何注记信息，那么势必会大大削弱地图的指示功能。因此，地图注记配置历来都是地图制作过程中一项必不可少的工作。注记位置配置的合理与否在很大程度上影响着地图的易读性，而注记配置速度的快慢也直接影响地图的成图周期。但是，从目前已有的商品化 GIS 软件来看，仍存在着注记配置位置不理想，速度慢的情况，尤其是注记的自动配置问题还没有很好地解决。从手工制图向计算机制图的转变过程中，地图注记的配置方法却没有根本性的变化，注记配置速度慢，过程繁复，最终效果不理想的情况是计算机制图中常见的现象。

②研究现状。许多科研机构在注记配置方面都投入了大量的研究并且相继推出了各自的注记配置系统。如美国人口调查局设计的注记定位系统，美国的 J. S Doerchler、H. Freeman 研制的大密度地名配置专家系统，美国纽约大学地理信息与遥感实验室在并行计算机上对地名和属性信息的自动配置进行的研究等。其中，人口调查局设计的注记定位系统已成功地用于人口普查。实践证明，它可以大大提高注记配置的效率。

在现有的一些流行的商品化 GIS 软件中，如 MapInfo，Arc/Info 也都有灵活实用、颇具特色的注记配置功能。如在 MapInfo 中，注记按其对应的地物分为点注记、线注记、面注记和说明性信息分别进行配置，配置的方法采用交互的方式，注记的各种参数如注记内容、尺寸、字体、旋转角度、位置等都以对话框的形式提供给用户并允许进行实时的设定和修改。

美国 ESRI 公司最新推出的地图注记自动处理软件 Maplex，用来快速生成各种比例尺的全符号化地图。在 Maplex 中用户可选择 96 种字体，支持 512 个字符的扩展符号集。注记配置的位置，对于点状符号有居中或偏移标注；对于线状符号 Maplex 沿着水

平或弧线方向放置线要素的注记；对于面状符号 Maplex 提供了若干配置位置的选择，有中央水平位置、中央曲线位置、中央连续位置和偏移位置。

Windows 操作系统的广泛使用已给汉字的显示与输出带来了极大的方便。但是汉字本身的许多特点也决定了在处理上与西文存在着较大的差异，而且我国的地图制作对注记的配置也有特殊的要求，这就给汉字注记配置的实现增加了不小的难度。已有的注记配置软件在处理汉字注记时，也有其力不从心之处。例如，Windows 提供的字库中并没有各种字体的变形字，而要将这些变形字扩充到现有字库中就需要首先了解字库的结构，进而对其编程，这往往是一个很费时费力的工作。这就给地图中常见的斜体字、耸肩字的显示提出了难题。自动注记配置过程实质上是一个复杂的人脑思维判断的过程。从这个意义上讲，它应属于智能化方面的问题。而现在需要把这一过程借助于计算机模拟出来，显然具有相当大的难度。

自动注记配置的关键难点和突破点就在于如何才能更快、更合理地自动寻找到每个注记的最佳配置位置。最近的研究结果表明，寻找最佳注记位置的问题是一个 NP－难度问题(NP－complete)，而且 NP－难度问题又是不可能用多项式时间算法来求解的。另一方面，在具体实现注记配置时，常会因为地图上某些地区的注记密度过大而造成注记位置回溯的现象，即在发生注记相互压盖的情况下对已经配置好了的注记重新进行定位，而且每次回溯的深度也不定。这样，要想在较短的时间内找到最优的注记位置从时间上来讲几乎是不可能的。

(2)交互式地图注记配置。交互式地图注记配置方式是目前普遍采用的一种注记配置方式。交互式注记配置的基本原理是注记的各种参数值(包括注记字的尺寸、间隔、色彩、旋转角度、字体、字形等)可通过人机交互的方式进行设定，注记的配置位置可以通过鼠标或键盘的拖动任意移动直到满意为止的一种注记配置方式。

这种注记配置实现过程不是不能采用，但速度太慢。有些 GIS 软件在这方面做了一些有益的改进。例如，先利用空间目标查找到注记，采用缺省的注记参数和一个缺省的位置（如目标的右边）实现该注记的配置；或是对一批具有相同属性的注记（如所有的村庄级居民地）采用一组参数统一进行配置。这样做不仅明确了地物与注记的对应关系，而且加快了注记配置速度。所以交互式地图注记配置的设计原则应首先从用户的角度出发，看如何才能使操作更方便、更灵活、更省时。

①交互式注记配置对数据的要求。交互式注记配置对数据的要求比较简单。这里提出三种途径来组织数据。

第一种情况，对于象地名一类的文本信息，它们可以从数据库或数据文件中直接得到。一般来说，一幅地图经数字化以后都会相应建立一个单独的地名文件。我们所要做的工作只是根据注记配置的要求，将地名文件中相关的注记信息提取出来。

第二种情况，对于这样一类注记，如桥梁的桥长、桥宽、载重吨数，水库的面积、容量，道路的路面宽、铺面宽、路面性质等数字信息以及数字和文字相混合信息的注记，配置时除了需要获得这些信息以外，还需要按照图式的规范要求，将它们组织成一定的书写格式。这种类型的信息在文件中一般是以描述码和参数码的形式存储的。描述码可以有多个，其对应值都是数字，要将这些数字解释为文字还需要对照属性编码表逐个要素进行翻译；参数码也可以有多个，而且一个参数码又可能同时表示两个以上的信息。那么如何提取这些信息呢？可以根据编码方案分别建立反映描述码、参数码与各类地物属性的对应关系的索引表。有了这些索引表，就可以将各种码值转换成相应的属性值了。

第三种情况，在一幅地图上，有时我们可能需要临时添加一些说明性的注记信息，对于这类注记可通过增加交互式的输入功能来完成。

②交互式地图注记配置功能模块。交互式地图注记配置功能

模块应该包括以下几个方面:注记信息的获取与显示,注记的交互式配置,注记的修改,注记的存取等。

——注记信息的获取与显示。在进行注记配置时,用户首先关心的是所要配置的注记究竟与哪一个地物目标对应,或是当前地物目标所要配置的注记内容究竟是什么,即用户希望自己在用鼠标点击某一地物目标时,该地物的注记内容就会立刻显示在旁边。这在具体实现时并不难,只是一个查找离鼠标点击位置最近的一个地物目标的算法问题。

——注记交互配置。当地物目标找到后,就可以进行注记的交互配置。那么注记配置在什么位置呢?这个位置一般是缺省给定的。如果是点状地物,配置位置可选在该点状地物定位点的右侧;如果是线状地物或面状地物,鼠标位置即可作为注记缺省的配置位置。

注记的其他参数(如注记尺寸、间隔、旋转角度、字体、字形、字色彩等)一般也是先给定的缺省值。地图上的地物多种多样,每一类地物注记的参数也各不相同。例如,同是居民地,但由于行政级别不同,注记的字体、尺寸也不同。所以应该根据各比例尺地形图图式规范先建立若干个索引文件,文件的基本格式可以如表 9—1 所示:

表 9—1 注记参数的索引

主码	识别码个数(n)					
识别码 1	字体	字型	尺寸(横、纵)	间隔(横、纵)	旋转角度	字色(前景、背景)
识别码 2	字体	字型	尺寸(横、纵)	间隔(横、纵)	旋转角度	字色(前景、背景)
………						
识别码 n	字体	字型	尺寸(横、纵)	间隔(横、纵)	旋转角度	字色(前景、背景)

其中主码说明类别，识别码说明级别。例如，主码可以是居民地，识别码可以是市、镇、乡、村庄等。

有了以上的索引表，就可以在获取地物(识别码)后，打开相应的索引文件获得注记的各参数值。至此，注记的全部信息都已得到，接下来要做的就是在缺省的位置，采用缺省的各注记参数，显示注记。

——注记修改。注记的修改功能应包括用户对注记的配置位置及其他各参数的修改。例如，用户对某个注记当前的配置位置不满意，就可以利用鼠标来移动该注记，直到满意为止。又如，用户对某注记的字体、尺寸不满意，也允许进行多次修改。对注记参数的修改可以通过提供对话框的形式来实现。当用户选定要进行修改的注记后，系统可以打开一个“注记参数对话框”，在对话框中显示的内容就是当前所选注记的各参数值。用户可以修改对话框中的参数值，一旦用户确认以后，系统就会将当前对话框中的各参数值赋给被修改注记并重新显示该注记。

——注记存取。注记经过以上交互式的配置和反复修改调整后，用户如果认为结果令人满意，就可以将全部的注记信息存储起来。注记的存储格式可以为表 9—2 所示：

这样，当我们打开一个已存在的注记文件，读取每条记录并输入到计算机当中，之后再进行的所有有关注记的显示、修改、输出等操作就非常方便了。

——交互式注记配置的辅助功能。交互式注记配置模块所应具备的主要功能是以上四点。这些功能在具体实现时，还需要根据实际要求添加一些有效的辅助功能。例如，通过对汉字注记特点的研究可以发现，在注记配置过程中，特别是在进行线、面注记配置时，通常希望能够将注记名称拆解成单个的汉字进行配置。这也是现在有些商品化 GIS 工具软件中所不具备的功能。针对这一要求，可以添加对注记内容逐字拆解的功能，使得用户可以对汉字串中任意的一个汉字进行操作。再比如，进行注记修改时，有时用户

表 9—2 注记的存储格式

字段名	类型	说明
ID	长整型	对象标识码。本表唯一,也作关键字。
Type	整型	注记类型 0:点 1:线 2:面 3:复杂 4:无对应地物
MainName	字符串	注记名
SubName	字符串	副名
Namement	整型	配置方式 —1:数字化的位置 0—7:点的八个方向 8—10:线的三个方向 11—13:面的三种方式
PointArray	变长实数组	注记点坐标串
FontType	整型	字型(扁,长,斜,耸……)
Font	整型	字体(细、粗等线,仿宋,宋,隶……)
Size	实型	注记尺寸
Direction	实型	注记方向(角度)
Interval	实型	注记间隔
Slope	实型	注记倾斜(角度)
ForeColor	长整型	注记前景色
BackColor	长整型	注记背景色
Effect	整型	注记效果

希望能够对某一类注记进行统一的修改,如改变所有村庄级居民地注记的字体,对所有面状要素注记改变字形等。这时,若只能逐个注记修改,显然工作量太大,太费时间。针对这一要求,可以添加注记的批处理功能,既允许同一地物类型注记的批量修改,以及允许同一类型注记(点注记、线注记、面注记)的批量修改,从而大大加快速度。

(3)地图注记自动配置。自动地图注记配置是指注记的定位位置以及各种参数都是通过索引表和算法自动确定的一种注记配置方式。其中注记定位位置的确定是地图注记自动配置实现的核心。地图上的注记按照它所说明的对象的性质不同,可以分为四类:点注记、线注记、面注记和说明性注记。前三种注记分别与点状地物、

线状地物、面状地物相对应，配置时要求注记对地物具有明确的指示作用，而后一种注记则不与任何具体的地物相关，位置由用户自由设定。这里主要是针对前三种注记进行自动注记配置的研究。

①点状要素注记自动配置。点状要素注记自动配置是目前注记自动配置中研究最多、最深入的一个方面，这主要是因为一方面地图上的点注记在数量上占了很大的比重，另一方面点注记的自动配置算法比起线、面注记来说也简单得多，而且点注记的研究成果也可以用于线、面注记。

在配置点注记时，一般应遵循如下原则：

——注记配置位置一般应位于图形或符号的右边或上边，其次是下边或左边；

——注记不能压盖点地物符号；

——注记不能压盖其他注记；

——可以压盖非同色线状地物，不能压盖同色且级别较高的线状地物，与同色且级别低的线状地物压盖时，该线状地物符号在压盖处应该断开。

——注记一般不压盖境界线，并且不超过其所在区域界限。

点注记自动定位大致可分为两步：

第一步：为待配置的注记提供一系列候选位置，该候选位置一般位于点状符号的外缘一定距离内。

第二步：综合考察各候选位置，选定较合理的位置作为当前配置位置。

在各种点注记自动定位算法中，候选位置一般都选定为八个方向，即以该点要素为中心点的八个邻域方向。这主要是出于两个方面的考虑：一是这八个方向基本上涵盖了点注记经常配置的位置。二是这时的注记配置速度不会太慢。

在具体配置注记时，哪个位置先选，哪个位置次之，哪个位置最后，必然存在一定的优先级顺序。根据点注记配置的原则，优先级顺序如图 9—19。

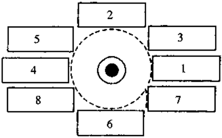


图 9—19　点注记自动配置的候选位置

对注记压盖的判断和处理是点注记配置的重点和难点。目前在已有的算法中几乎都是将注记看作多边形(一般为矩形),这样,在自动配置过程中,解决注记不压盖点地物符号或注记间不相互压盖的问题就转变成了解决多边形之间互不相交的几何问题了。同理,注记不与某些线状符号压盖也就是多边形与线段不相交。

在实际配置注记时,如果注记的密度比较大,就很可能出现注记间相压盖的现象。这时,需要判断是否可以通过移动注记的位置,解决压盖现象,最终实现注记的合理配置。有时这种移位不仅针对当前注记甚至还会涉及到已配置的若干个注记。我们把这个过程称之为注记位置的"回溯"。简单地讲,回溯就是注记出现压盖时,对压盖的注记进行重新定位,从而避免压盖的过程。注记回溯是自动注记配置中最重要的一个环节。

那么根据以上提出的八个候选位置,如何才能确定一个最合适的注记配置位置呢?我们可以通过采用将其中一些环节量化的方法来近似模拟以上过程。首先把注记在某位置配置时可能影响其配置合理程度的因素都总结出来,然后按照主、次赋于不同的权值,影响程度大的权值就大,反之,权值就小,并将各权值相加,得到的和就作为该注记配置在此位置的综合指标。同样的步骤计算出八个位置上的权值和,最后选择权值和最小的那个位置作为注记的当前配置位置。

不难看出，以上的描述并没有考虑到回溯，这样就不能保证比较后得到的权值和最小。以图 9—20 为例，假设只考虑位置和压盖两项权值，位置权值按照注记位置的优先级顺序从 1——8；有压盖时，压盖权为 20，无压盖，压盖权值为 0。地物 A 的注记已配置完毕，地物 B 的注记配置在八个位置上((a)——(h))分别得到一个权值和(f B)。

比较(a)到(h)8 个值可知，位置(d)因为权值最小(f B＝5)而被认为是当前最佳位置，但实际上如果对 A 位置稍加调整就可得到 (i)情况，这时的权值(f B＝3)显然比选择位置(d)时的权值小，因此位置(i)比位置(d)更佳。

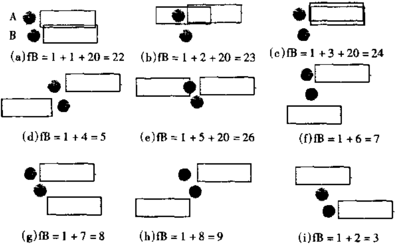


图 9—20 点注记各方向的目标函数值

再如图 9—21，假设 A、B、C、D、E、F 注记都已配置完毕。对于 G，如果不考虑注记回溯的移位，那么 G 将没有位置可进行注记。但是假设这时只有 F 注记可以移动到右上方，那么只要将 F 移动到该位置，G 注记就有地方配置了。由此可见，注记位置的回溯问题在注记自动配置中是必须考虑的。

计算注记位置权值时，目标函数的最终值可以表示为：目标函

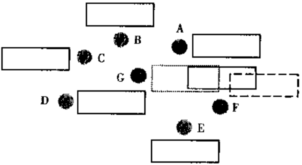


图 9—21 注记位置的回溯

数值＝位置权＋点符压盖权＋注记压盖权＋线符压盖权。其中位置权是从 1 到 8 的数值，点符压盖权和注记压盖权应是一个较大的数值(如 20)，线符压盖权也是一个较大的数值，但最好小于点符压盖权和注记压盖权，因为注记可以压盖某些线状符号。线符压盖权的权值大小也不见得都相等，可兼顾所压盖线符的性质(如等级、面积、数量等)赋不同的值。如果压盖的线符等级高则权值大一些；等级低，权值就小一些。压盖线符的面积大，权值大；反之，权值就小。压盖线符的数量多，权值就大；反之，权值就小。

点注记自动配置的步骤如下：

· 获取点状要素的定位点坐标，确定八个候选位置；

· 计算每个候选位置上的目标函数值；

· 判断每个位置有无注记压盖，若有，进一步判断是否能够将被压盖注记移动到另一个位置后可解决压盖问题，若能，则重新计算注记压盖权和目标函数值；

· 选取权值最小的位置作为该注记的定位位置；

· 配置当前注记，若存在移位，还要将涉及到移位的所有注记重新配置到新位置上；

· 判断是否点注记全部配置完毕，若没有，回到第一步，否则结束。

②线状要素注记自动配置。在配置线注记时，一般应遵循如下

原则：

——注记配置位置可以是沿线或平行线配置，平行线配置时又可以分为沿上平行线和沿下平行线配置；

——注记不能压盖点状地物符号；

——注记不能压盖其他注记；

——根据制图规范，线状要素长度超过一定值时，要分段重复注记。

线注记自动配置算法的基本思路是：首先判断注记是沿上平行线还是沿下平行线配置，其次按注记字个数等分线状要素，确定注记的初步定位点并判断这些位置是否有与点符或其他注记压盖的现象，若有，进行移位，直到找到合适的位置为止。移位时，在不调整字距的情况下，先在现有位置上前后移动一定距离（设为两个注记字长度），直到与点符不再压盖，如果不行，再上下移动，如果还不行，再调整字距。

按照以上的基本思路，当注记中有一个汉字不能配置时，其他汉字都要重新确定位置，当重复注记时，只要有一段注记位置调整了，其他段的注记也要重新调整位置。显然这样的配置过程太繁复，而且效果也不一定好。可以对这种算法做以下改进：

——先进行线要素注记的配置，从而避免了线注记因压盖而造成的移位问题。

——线注记的数量一般不多，那么它与点符压盖的情况自然很少，因此即便个别压盖了，可以通过注记交互修改功能进行位置调整，其目的就在于简化线注记配置过程、节省时间、提高速度。

——为了保证注记位置适中，可以首先将线状要素总长度进行首尾的定长截取（如图 9—22），得到 a、b 两点。然后计算 a、b 两点之间的线状要素长度，按照注记字的个数（N）平均分成 N－1 份，则等分点就是该线注记的注记定位点。

根据线注记自动配置基本算法以及所做的改进，线注记自动配置的步骤如下：

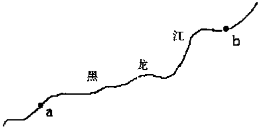


图 9—22 线状注记配置

——提取完整的线状要素，判断是否需要分段注记；

——对每段进行按注记字的等分，找到等分点；

——依次在等分点上配置注记；

——如果需要分段注记则回到第二步，否则结束。

在各种线状要素的注记自动配置中，等高线高程注记的自动配置比较特殊，其基本方法可描述如下：

首先必须知道等高线的高程值，然后把需要进行等高线注记的区域划分为若干个注记空间，只允许在一个注记空间内配置一个高程注记，这样可以控制注记配置的密度和使注记均匀分布。

在一条等高线上进行注记时，要顺着等高线的走向去找合适的注记位置，然后进行具体的注记配置。所谓合适的注记位置是指：

——处在一个没有配置注记的注记空间中；

——应在较平直的等高线上注记，这时要求计算等高线的曲率；

——不应在等高线较密的地方进行注记，这可以通过注记位置上有多少条等高线通过来判断；

——有时要求不能压盖地形特征线，这时如果有地形特征线的坐标，就可以很方便地进行判断。

对于等高线的注记，要求字头朝向高处，注记字平行于等高线。在找到了等高线注记的合适位置后，可以通过该条等高线在附

近点的平面坐标算出注记的斜率。注记字头朝向的判断可以通过判断该条等高线两边的高程来确定。这时就可以进行具体的等高线高程的注记配置了。对等高线进行高程注记配置就是写数字串，只要对其进行适当的旋转、平移、缩放就可以了。

等高线在注记处必须断开。为了做到这一点，可以先对所有等高线进行注记配置，等到所有的注记全部生成后，把它记录在文件中，形成一个注记文件。要绘等高线时，判断等高线是否压着注记。如果压着注记则必须断开，在离开注记时再继续绘，这样就可以保证等高线不与注记重叠。

③面状要素注记自动配置。在配置面注记时，应遵循如下原则：

——面状要素注记应尽量配置在面域内部，只有当面状要素面积小于一定值时，才将其注记放置在面域之外。

——配置在面域之外的注记不能压盖重要地物、境界线和已有注记。

——面积大于一定值的面状要素注记要重复配置。

——注记配置在面域内时，一般沿水平（或垂直）方向或沿面要素的主骨架线配置。

——当注记在面域内配置不下时，可以适当缩小注记的尺寸。

由于地图上面状要素所对应的几何图形都是多边形，因此，对面注记自动配置算法的研究也就集中在了对多边形的研究上，其中所要解决的关键也是难点问题就是求多边形的主骨架线。所谓骨架线简单说就是多边形的走向线。可定义为到多边形的两个或更多点的距离相等的点的集合。主骨架线则是指骨架上一条最长的连续线。如图 9—23 所示。

图中短虚线是次骨架线，长虚线是主骨架线。可见只要求出多边形的主骨架线，将注记沿主骨架线均匀配置，就能实现面注记的自动配置。但是，由于地面上面状要素的形状多种多样，要提取出合理的主骨架线还是相当困难的，这无疑就增加了面注记自动配

置的难度。



图 9—23 面要素的骨架线

目前面状要素注记自动配置归纳起来主要包括以下几个步骤：

第一步：提取多边形，化简多边形。

第二步：提取多边形主骨架线。

第三步：沿主骨架线均匀配置注记。

多边形骨架线提取算法的基本思路是：首先根据多边形的主走向（横向或纵向）决定对多边形进行水平切割还是垂直切割，然后对多边形进行修整，再计算每条切割线与多边形的交点，最后求出每组交点的中点，将所有中点连线就得到了该多边形的主骨架线。

使用此算法的一个条件就是进行水平或垂直切割时，每条切割线与多边形的交点要保证有且不能超过两个。这是因为若允许交点多于两个，就不能确定到底取哪两个点的中点坐标作为骨架线上的点。如图 9—24 中的 1，2，3 点在一条直线上，难以确定取 1，2 还是 2，3 之间的中点作为主骨架线上的点。

仔细分析可以知道产生这种情况的主要原因是多边形含有"极点"。极点是指该点的（ X，Y ）值同时小于或大于其前驱点和后继点的（ X，Y ）值。图 9—24 中的 2 点就是极点。因此在对多边形进行水平或垂直切割之前必须先对其化简，目的就在于去掉极点，从而避免求骨架线出现"二义性"甚至"多义性"。去除极点的方法是：直接在极点上进行水平或垂直切割、然后求交并构成子面积 S_1 和 S_2。当 S_1、S_2 都小于一定值时，就同时舍去 S1、S2，否则保留

面积较大的多边形。

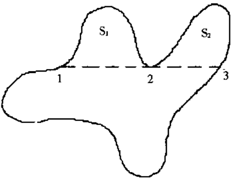


图 9—24　化简多边形，去除极点

仔细研究地图上的面注记，会发现很大一部分面注记是沿直线(或经纬线)方向配置的。因此对面状要素可以先判断是否能按照水平或某一角度进行配置，如果不行，再沿骨架线配置。其具体的算法可描述为：

首先按照注记的尺寸计算注记所占的矩形框的大小，再求出多边形的中心点。注记以多边形的中心点为中心从 0°～360°按一定间隔旋转，当旋转到某一角度时矩形框完全落在多边形内，记住此时的角度 α_1，继续旋转直到下一个临界值 α_2，当角度在 α_1 与 α_2 之间时，矩形框完全落在多边形内，这时求 α_1、α_2 之和的平均值作为注记的配置位置。

根据面注记自动配置的基本算法以及所做的改进，面注记自动配置的步骤如下：

・提取面要素的几何坐标；

・判断注记是否可以沿水平或某一角度实现配置；

・如上一步不能实现，对多边形进行化简(包括去掉小弯曲和极点)；

・提取骨架线；

・沿骨架线均匀配置注记；

・面注记是否配置完毕，否，转到第一步；是，结束。

在进行面状要素注记自动配置时，若面要素的面积小于一定值，就需要将注记配置在面域之外，这种情况类似于点状要素的注记配置，其方法可参照点注记自动配置的算法。

对于地图注记的自动配置，有一点我们必须认识到：自动注记配置并不是实现注记配置的完全自动化，不合理的现象依然存在，交互式的修改功能必不可少。因此，从这个意义上讲，“自动注记”只是一定程度上的“自动”，自动注记配置系统必须要有人机交互的辅助功能。

(4)屏幕地图中地图注记的显示。对于GIS中的屏幕地图，用户可以很方便地进行放大或缩小，这时屏幕地图上的注记内容也要随之变化。这种变化通常分为两种形式，一种是注记随着屏幕地图的放大、缩小而放大、缩小，另一种是注记尺寸不随屏幕地图的放大、缩小面变化，但注记内容却在变化。

对于第一种情况，大多是把屏幕地图当作硬拷贝地图(如纸质地图)来使用时出现的情况；当在屏幕上检查即将输出硬拷贝地图的结果时，为了所见即所得，即为了使在屏幕上出现的情况与输出后出现的情况完全一致，也采用这种随屏幕地图的比例尺自动缩放的方法。

为了使注记的尺寸随着屏幕地图的比例尺而变化，只要把注记的尺寸转换到原始数据的比例尺上，则当给定新的比例尺时，可算出在新比例尺下注记的尺寸，这样就可达到随着地图比例尺的缩放，注记也作相应缩放的目的。

设在进行地图注记时，屏幕地图对原始数据的放大或缩小倍数为 P_0，而这时注记的尺寸为 S_0，则和原始数据有相同比例尺的注记的尺寸 Ss 为：

$$S_s = S_0 / P_0 \qquad (9-1)$$

可把 S_s 作为注记字的原始尺寸记录在注记文件中。设对制图区域进行了开窗放大(或缩小)后，屏幕地图对原始数据的放大(或

缩小)倍数为 P_1，则这时屏幕注记的尺寸为 S_1：

$$S_1 = S_s \cdot P_1 \tag{9—2}$$

根据尺寸 S_1 绘出地图注记时，就可使注记适应比例尺的变化。要注意对于单线体矢量汉字的粗细也应作相应的变化。

对于第二种情况，是在把屏幕地图当作电子地图使用时出现的情况。这时，注记的尺寸必须使用户能在屏幕上清楚地识别出每个注记。因此，当屏幕地图缩小时，为了使用户能看清楚注记，注记的尺寸不能改变，但为了避免注记之间的压盖，必须自动地舍弃一部分注记；而当屏幕地图放大时，自然就可以显示更多的注记了。所以，在这里注记的自动配置是必须具备的功能。

对于有些 GIS，它们具有多比例尺的空间数据。因此，当屏幕上的地图被放大或缩小时，地图上的内容会发生变化，也就是说，GIS 会根据屏幕地图的尺寸范围和比例尺，自动在 GIS 的地理数据库中选择合适比例尺的数据来显示。例如，在屏幕上显示中国全图时，GIS 会选择小比例尺的空间数据；当放大到一个省的范围时，GIS 会选择较大比例尺的空间数据；当放大到一个城市时，GIS 会选择更大比例尺的空间数据。在这种情况下，地图注记应能根据屏幕上所显示地图的数据情况，自动进行适当的配置。

2. 地图图例配置

正是有了地图图例，读者才可以快速而准确地从地图上获取信息，可以说图例是理解地图的钥匙。关于地图图例设计，通常提出两点要求：

- 图例必须完备，即应包括地图上所采用的所有符号；
- 图例中的符号与地图中使用的符号必须完全一致。

但对于屏幕地图，由于屏幕范围的限制，往往只表示专题要素的图例，其他图例可通过查询获得。

由于地图内容和表示方法各不相同，图例的内容和复杂性也各不相同，所以往往要用人机交互的手段来进行图例配置，过去甚至使用临时编程的方法。

地图图例的配置是在确定了地图要素的表示等级(定性表示、分类表示、顺序分级、间隔分级、数值表示)和相应的符号化方法之后进行的,地图图例就是描述地图要素及其表示等级与其相应符号之间的关系。因此,只要把前面数据处理和符号化的结果显示出来即可。例如对于粮食亩产分为五级:300 千克以下,300～400 千克;400～500 千克;500～600 千克;600 千克以上,用分级圆的方法表示时(这时已知道了每个等级符号的尺寸、色彩),地图图例描述的就是粮食亩产的每个等级与其对应的分级圆符号的一一对应关系。

把相应要素的符号和所表示的类型和等级表示出来是可以用软件实现的。例如对于用分级圆表示粮食亩产的情况,可以由软件自动地形成,如图 9—25。这只要根据符号各个等级的尺寸,以及指定图例中符号间的间隔,即可绘出图形;根据每个等级的符号所对应的数据(可从数据处理后的文件中获取)和量纲即可在旁边绘出符号对应的数值和量纲,有时也可采用人机交互注记配置的方法对图例中的注记进行配置。

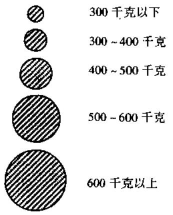


图 9—25　用分级圆表示分级的图例

这时虽然已绘出了地图图例,但图例的定位并没有确定。确定图例定位的方法通常有两种:一种是把图例当作一个图例块,用户

只能移动整个图例块在地图上找到适当位置后确认，如图 9—26 所示。但这种方法在制图区域内容较满或图例尺寸较大时往往找不到合适的位置。第二种方法是允许移动图例中的每个符号和注记，更灵活地配置注记，但操作比较麻烦。

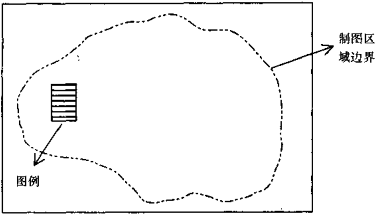


图 9—26 图例交互定位

在对面状符号设计图例时，通常要指定表示面状符号的矩形的大小，但对某些要素不能用矩形表示，例如对于湖泊，应绘出一个湖泊；同样对于线状符号，通常以直线段的符号来表示，但是有时不能以直线表示，如表示河流时，应表示弯曲的河流。对这些特殊情况，在图例配置软件设计时都要考虑到。当有许多图例设计方法时，如图 9—27 也是一种表示亩产分级的图例，这时可给出最常用的一种，或允许用户进行选择。

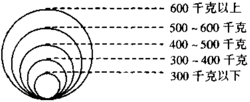


图 9—27 一种表示分级的图例

§9.4 GIS的自动制图综合

1. GIS自动制图综合的意义

(1)GIS自动制图综合的必要性。在传统的地图生产中,制图综合是制图人员最关心也是研究得最多的一个问题。这是由它在地图生产过程中的地位决定的。在数字地图制图条件下,数字地图的自动制图综合又成为所有数字地图制图人员普遍关注的问题。GIS的迅速发展和广泛应用,在一个更高的层次和更广泛的意义上把地理信息的自动制图综合问题摆在了我们面前,因为它直接影响到地理信息的深加工和地理信息产品的生产。

GIS中地理信息的自动制图综合,其目的在于建立一个多尺度或"独立于比例尺"的GIS,或者建立一种比例尺的地理数据库,而供多种比例尺的GIS显示、分析和制图输出。屏幕显示是GIS应用的重要方式,由于屏幕大小和屏幕(电子)地图阅读方式的特殊性,显示内容和区域范围必须是有选择的,这就要靠地理信息的自动制图综合来解决。基于GIS地理数据库的空间分析,可用于宏观和微观的不同目的,用以进行分析的数据的详细程度是不一样的,这也需要通过地理信息的自动制图综合来解决。至于GIS的制图输出就更不用说了,由较大比例尺的数据编制较小比例尺的地图时,只有通过地理信息的自动制图综合才能实现。

总而言之,GIS的地理数据库不可能存储每种比例尺的图形(图像)数据,需要解决如何由一种比例尺的地理数据通过综合以满足其他比例尺的数据显示、分析和制图输出的需求。这是一个不能回避的问题。

(2)GIS自动制图综合的两类任务。对于GIS的自动制图综合的问题,目前有两种看法:一是认为综合后的数据应是原始数据本质的、主要的和整体性的内容,而应舍去非本质的、次要的和局部的细节;二是认为GIS的综合原则不应当完全等同于一般的地图

制图综合，或者说应当区别于输出目视化地图产品的综合。人们认为在 GIS 的地理数据库中存储一种比例尺地图数据库，供多种比例尺 GIS 分析和制图输出使用，这是需要的，但是可以分为两类：一类是满足于 GIS 屏幕显示和分析需要的自动制图综合；另一类是满足于输出较小比例尺可视化地图产品(纸质地图)的自动制图综合。前者是地理数据的综合，一般情况下可以不进行关系处理，不必考虑地图生产规范的要求；后者是地图数据的制图综合，需符号化和进行关系处理，必须符合地图生产规范的要求。

(3)GIS 自动制图综合的特点。由上可以看出，GIS 自动制图综合需解决的问题，要比传统地图生产中的制图综合和基于地图数据库的数字地图的自动综合需解决的问题要广泛而深刻得多。由于地理信息要比地图信息更广泛，因此，地理信息的制图综合包含两个方面的内容：一是图形信息的自动综合，二是专题信息的自动综合。

2. 自动制图综合的基本方法

(1)图形信息的自动制图综合。和普通地图的制图综合一样，图形信息的自动制图综合主要包括地物轮廓形状的化简、各要素的选取、图形的合并、图形等级的变换、重要图形的夸大、要素相互间关系的处理等。

①轮廓形状的化简。轮廓形状的化简既要减少数字化线划要素或面状要素边界中一定数量的数据，同时又要保持该要素的图形特征。其大致过程是，先识别出特征点，再按照用户设定的指标去掉不必要的碎部。

有关数字线划图形的化简，是自动制图综合研究得最多，且效果最好的一个方面。美国学者 McMaster 将化简算法分为分散点算法、局部处理算法、不受限制扩大的局部处理算法、受一定条件限制扩大的局部处理算法、整体化简算法(如 Douglas－Peucker 算法)。虽然化简的方法很多，但常用且效果较好的还是 Douglas 算法。由于它本身还有其一定局限性，对于一些复杂的有特殊要求

的线划化简效果不够理想，所以在实验中又提出了带某种约束条件的 Douglas 算法（如化简海岸线时扩大陆地缩小海域）。

②要素的选取。当比例尺发生变化或用途发生改变时，应通过有选择地保留或删除一些要素，以达到清晰显示或输出的目的。要素的选取主要解决两个方面的问题，即选取多少和选取哪些。将定额选取模型、结构选取模型、选取过程算法、选取规则同 GIS 的各种检索功能有效地结合起来是实现要素选取的基本途径。

③图形的合并。比例尺缩小后，图形之间靠得很近以致无法逐个表示出来，这时就要把一些相互之间靠得很近的要素合并表示。最常见的、最典型的例子就是面状居民地街区之间的合并。

④图形等级变换（表示方法的改变）。主要指由于受比例尺的限制而不能以原来的图形等级表示，而必须改变其表示方法的那些要素，常见的有线状要素变为点状要素表示，面状要素变为线状要素表示，双线到单线等。

⑤夸大。即对要素的局部进行放大，以达到清晰和强调的目的。要素局部的夸大表示提高了视觉效果和要素特有的特征，例如，一个小的岛屿，如果仅从其大小来看或许应该删掉，但它若具有导航价值，则需要将其保留并夸大表示出来。

⑥关系处理。为了避免与重要要素的符号发生冲突，对次要要素和符号需要进行移位处理，移位处理前，必须首先建立要素的优先级别和移动顺序。

对于图形发生冲突的要素实施移位，其过程大致可分为以下几个步骤：检查图上是否有明显的图形重叠，计算位移的方向和大小，移开次要的要素，验证位移后的效果。

根据位移要素的性质，移位的可以是点状要素、线状要素，也可以是面状要素（全部轮廓线的移动或部分轮廓线的移动）。

利用上述各种图形信息的自动综合，可以形成一个自动制图综合的算子库，以便在综合时从中提取有关算子进行分项操作。当然，由于自动制图综合的复杂性，还必须有其它工具软件的支持，

如检索、查询、交互编辑等。

(2)专题信息的自动制图综合。对于专题信息而言,它的自动制图综合除了数量的选取、形状的化简以外,更重要的是现象的质量特征与数量特征的综合,或者说,它是由数据到数据的综合,其综合结果的体现是形成新的分类分级体系。

专题现象质量特征的综合表现在多个方面,但其根本任务是用概括的分类代替详细的分类,以整体的概念代替部分的概念,以达到数据综合的目的。

专题现象数量特征的综合,主要是以扩大分级间隔,减少分级数的方式进行的。

常用的专题数据的分类分级方法有:

——任意分级或逐渐增大的分级,适用于有实际意义的重要分界点的数据。

——按数据的分布特征进行分级,重要的是采用科学的方法观察数据的分布特征。

——自然聚类法,将全部数据按其增长顺序排列,计算每两个相邻数据之差,并将之与聚合半径相比较,据此进行分级。

——系统聚类法,对于所有数据建立原始数据矩阵,对其标准化后建立相似矩阵,然后利用相似矩阵进行聚类分级。

——最优分割法,是在有序样本不被破坏的前提下,其类距离用离差平方和表示,并使其分割的类内离差平方和达到极小,而类间离差平方和达到最大的一种聚类方法。可以按误差函数公式计算分级误差的大小,以择其优。

专题信息自动制图综合的本质,是利用数据分析的方法,对地理信息进行深加工,直接为用户提供结论性专题数据集和专题地图,以达到自动制图综合的目的。

(3)综合结果的显示与输出。地理信息制图综合结果的屏幕显示,所涉及的只是地理数据综合结果的显示,并不需要严格遵守地图生产规范的要求,也可不必考虑关系处理问题,如果目视可读性

受影响的话，可采用开窗放大的形式使其清晰化。而对于“硬地图”的制图综合来说则不然，它必须遵守地图生产规范的要求，必须使输出图形之间的关系正确，必须有相应的符号库的支持，以保证输出地图产品同常规地图生产的产品完全一致。

3. 自动制图综合的发展趋势

自动制图综合的发展趋势包括以下三个方面：

(1)智能化系统的开发。目前，计算机系统作为工具正在越来越有效地取代烦琐的手工劳动，但另一方面，人们又不能不承认，有时一个简单的人工干预动作，计算机通过无数次的迭代或递归还未必能圆满完成，综合中难于实现自动化的方面并不完全是因为具体的要素选取、化简、夸大、合并、移位等算法或规则的选择，而是如何让计算机学会随机应变。目前的制图综合系统之所以没有应变能力，主要原因之一是人们还不能把专家知识用计算机语言准确而全面地表达出来。

专家知识的采集，除了必须了解手工综合的机制，征求专家的意见，整理有关资料和文献外，还可以用智能增强法、反向工程法和差别适应法来收集专家的知识。

专家知识的计算机表达，可以是基于知识的系统，也可以是神经网络系统。基于知识的系统是人工智能领域应用最广泛的一个分支，其最基本的组成部分是数据库、知识库和推理机。基于知识的专家系统之所以至今还不能充分发挥其作用，主要是由知识库的不完整和知识表达的不充分造成的。

神经网络是由大量处理单元广泛互连而成的网络，是在现代神经科学研究成果基础上提出的对人脑功能基本特征的某种抽象、简化与模拟。神经网络具有学习能力，它是一种适合于并行处理的数学模型。较之串行计算机系统，神经网络具有更大的结构稳定性和抗干扰能力。当某一部分神经元受到阻碍或被破坏时，并不会给整体行为造成致命的影响。而且神经网络的训练过程实际上同时隐含了手工制图综合知识的采集与表达，这无疑为地理信息

的自动制图综合提供了一个新的思路。

(2)分形方法在综合中的应用。分形几何是一门专门研究非线性复杂系统的新型理论。分形是指一类具有相似结构的复杂几何形体，其核心是自相似性，其特征量是分数维，其突出特点是无特征尺度。自相似性简单地说就是局部与局部或局部与整体之间在形态、功能、信息、时间或空间等方面具有相似性。由于制图现象中大量存在统计意义上的自相似现象，故可以采用分形的方法进行综合。

分数维的确定可有多种定义和计算方法。对于综合的对象，可以通过计算分数维来量化其结构特征。一般地说，曲线的分数维可以是 1～2 之间的任何值，曲面的分数维可以是 2～3 之间的任何值，分数维值越高，图形就越复杂。

分形几何用于自动综合的重要意义在于，它可以为综合对象的形状和复杂程度提供量化指标，从而为图形的简化提供数学模型。

(3)现有方法的可行性检验及综合结果的质量控制。为了减少不必要的重复研究，现有的自动制图综合方法应得到合理的利用。由于每种方法都有它的适用范围和约束条件，因此在不同的场合会产生不同的效果，所以对现有方法的可行性分析及检验不仅能够保证综合的质量，而且可以避免引起盲目的指责和无意义的改造。

对数据进行综合比在传统地图载体上进行综合要快捷得多，这一方面显示了计算机的优越性，另一方面也意味着数据被歪曲的可能性增加了。数据的歪曲增加了相应的可视化产品的欺骗性，因此，对计算机制图综合结果的质量控制就显得尤为重要，这要通过一系列的判断与校正才能得以实现。

§9.5 GIS中地图设计的智能化

1. GIS中的地图设计问题

GIS是对空间信息进行采集、存贮、管理、分析、显示和输出的计算机技术系统，对于用户来说，最重要的是空间信息的显示、查询、分析和制图，而空间信息的显示、查询和分析都是以地图作为媒介的，GIS中的制图功能就是直接利用GIS中的地理数据来制作地图。由此可见，地图既是空间查询和分析的基本媒介，也是GIS的主要产品。

随着GIS应用领域的扩展，几乎任何人都可能利用GIS制作地图(包括硬拷贝的纸质地图和软拷贝的电子地图等)。为了制作地图的方便，GIS提供了图形编辑、注记配置等功能，且允许用户选择色彩、创作符号等，这样设计出来的地图可能色彩鲜艳，但很可能不能有效地传输信息，有时甚至会产生误解。这种情况的出现不是由于GIS软件不能制作地图，而是由于GIS的用户没有意识到所形成的产品是地图，而地图则意味着要传输信息，要有效地传输信息就必须使所生成的地图符合地图学基本原理。这就是说，要使GIS生产出能有效地传输信息的地图，就必须要经过地图设计过程。

地图设计工作往往需由那些具有丰富的地图设计知识和经验的专家来完成，要大面积地提高GIS用户的地图设计水平几乎是不可能的，因此，必须在GIS中加入地图设计专家系统模块，以便使GIS的用户在没有地图设计专家知识的情况下，也能进行具有专家水平的地图设计工作。

2. GIS中地图设计专家系统的任务和结构

地图设计专家系统的任务是根据地图设计的前提条件，确定地图的数学基础、内容和表示方法等。GIS可根据地图设计专家系统的推理结论，再经过适当的整饰和处理，即可显示和输出地图。

地图设计虽然是一项非常复杂的工作，但通过分析可以发现，地图设计工作可以划分为具有一定顺序的任务块，且这些任务块所涉及的知识也是相对独立的。因而可以通过针对不同任务块来建立相应的子专家系统来简化地图设计专家系统的复杂性，使之便于实现和维护。这些任务块的功能可描述如下：

——确定地图设计的前提条件。这是指进行地图设计时，需要知道所设计地图的制图区域；地图的类型，是电子地图还是纸张地图，是用于显示、查询、分析、还是输出；地图的主题等等。这些信息中有的可以从 GIS 的操作中获取，例如在对 GIS 的分析结果制图时，就已知了其制图区域、主要的地图主题等，而有的需要向用户询问，例如在直接利用 GIS 地理数据制作地图时，就需要向用户询问制图区域、地图主题等问题。

——确定地图的数学基础。这里主要是指投影和比例尺的确定。地图投影可以根据制图的目的、制图区域的形状等信息推理出来，比例尺则可由图幅大小和制图区域大小来确定。比例尺是控制地图数据的精度和地图内容详细程度的指标之一。

——确定地图内容。这是指专家系统应推理出地图上应表示哪些地图内容，这样才能进行下一步的地图设计工作。地图内容的确定可以根据地图设计的前提条件和地理数据库的内容推理出来。

——确定地图内容的表示程度。这有两方面的含义：一是当地理数据库中有不同精度的数据时，应选择合适精度的数据来制图，当数据的精度不能满足要求时，应给出提示；二是指对专题数据的分类分级，例如，对于某一项地图内容，应确定是否分类分级，如何分类分级等。

——地图内容的数据处理。这是指根据一定要求，对空间数据进行一定的制图综合，对属性数据选择适当的数字模型进行处理（如分类分级等），以便使所设计的地图能更有效地传输信息。

——地图内容的符号化。这是地图设计专家系统的核心内容，

即应推理出每一项地图内容所使用的符号，不仅要确定每个符号的形状、色彩、尺寸、图案等视觉变量，而且还要保持地图符号间的总体协调。

——地图的注记配置和图例配置。这是指应根据地图的用途、地图的内容等信息，确定地图注记的字体、尺寸、色彩、定位点等；地图图例是描述地图内容与其符号之间关系的，而地图设计专家系统不仅确定了地图内容，而且确定了地图内容的符号化方式，因而就可以确定地图图例的内容。

当地图设计专家系统的各子系统顺序完成了以上的任务后，就得出了 GIS 中所需的地图设计结论，也就是说 GIS 知道了从地理数据库中取出什么数据、对数据进行什么处理、用什么符号来表示地图数据、每个符号的形状、色彩、尺寸等是怎样的、注记如何配置、图例应具有什么内容等，换句话说，GIS 就可以运用这些结论来显示和绘制地图了。当然，要正式出图还需进行一些图形整饰等编辑工作，而这些是 GIS 原来就有的功能。

作为 GIS 中的地图设计专家系统，不能象常规的专家系统一样经常给用户提问题，因为地图设计中的许多问题是非专业人员很难回答的，所以应尽量用默认或缺省的方法来处理可问可不问的问题，对必须询问的问题，也必须保证普通的 GIS 用户能回答。当用户对专家系统设计的结论不满意时，也应该允许使用 GIS 中的图形编辑等功能对推理结论进行修改，因为对地图设计的评价还没有统一的标准。

3. 地图设计专家系统与 GIS 的连接

GIS 中的地图设计专家系统需要为空间数据的显示、查询和分析设计基础底图，需要为把空间分析的结果制成地图而进行地图设计，也需要为直接利用地理数据库生产地图而进行地图设计，因此，具有地图设计专家系统的 GIS 基本结构由图 9—28 所示。

其中，知识库存储地图设计所需的各种知识，模型库除了要存储空间分析模型外，还要存储专家系统所需要的各种模型（如分类

分级模型等)，而符号库的建立、管理和使用应满足地图设计专家系统的要求，即把组成 GIS 中的地图符号的各种参数与地图符号的视觉变量联系起来。

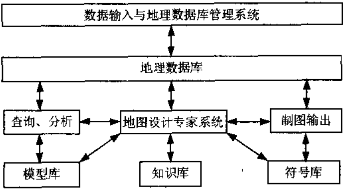


图 9 28 具有地图设计专家系统的 GIS 基本结构

为了使 GIS 与地图设计专家系统能顺利连接还应考虑以下问题：

——地图设计专家系统如何从 GIS 中获取有关的操作信息。例如，用户正进行的空间数据的显示、查询、分析等操作可作为地图设计专家系统的前提条件之一，因为它影响到要显示地图的目的；再如空间分析的对象、要求、范围、窗口等信息，也可为空间分析结果的制图输出提供某些前提条件。这样就可在尽可能不向用户询问的条件下进行专家系统的推理。因此，当在 GIS 中加入地图设计专家系统模块后，应显式地把有关操作信息记录下来，以备地图设计专家系统使用。

——地图设计专家系统如何与 GIS 的地理数据库连接。专家系统不仅需要对地理数据库进行实时的数据提取操作，而且还需要知道地理数据库中的数据质量信息，如数据的来源、时间、误差等信息，以便于提取合适的数据，并对数据进行适当的处理。而目前 GIS 中关于数据质量的信息还非常缺乏。

——如何使GIS的符号库能满足地图设计专家系统的要求。目前GIS的符号库是以显示和绘制为目的的，没有考虑地图符号设计的要求。在地图设计专家系统中，是根据地图要素的特征和地图符号的视觉变量来进行设计的，因此，需要把符号中构成符号的各种参数与地图符号的视觉变量有机地联系起来。

——如何使地图设计专家系统能自动地选择GIS中的数学模型。为了使地图设计专家系统更少地向用户提问，专家系统应能够根据所需处理数据的名称、类型、分布特征等信息，自动地选用合适的数学模型，这样GIS的模型库中应给出模型的特点、适用条件、适用范围等信息。

——GIS如何理解地图设计专家系统的推理结论。这需要对地图设计专家系统的推理结论的数据结构或知识框架进行标准化，同时，GIS的显示与绘图软件也应能理解地图设计专家系统的推理结论。

GIS的应用领域正在迅速扩展，作为空间数据的有效表达方式——地图，可能出自不同类型和层次的用户，而目前的GIS软件并不能使所设计出的地图符合地图学原理，因此有必要在GIS中加入地图设计专家系统模块，使GIS所产生的地图能有效地传输信息。

国内外有关地图设计专家系统的研究虽有不少，但基本上都是实验室产品，为了把地图设计专家系统模块加入到GIS中去，在地图设计专家系统和GIS方面都还有很多的研究工作要做。

§9.6 GIS的地图输出

地图输出就是将抽象的空间地理数据转换成可视化地图图形的过程，是空间地理数据可视化最常用、最有效的途径。空间信息的图形经符号化以后，输出的形式有两种：一种是屏幕显示输出，一种是硬拷贝输出。

1. GIS 中地图输出的特点

作为空间和空间关系的图形表示，GIS 中数字地图的基本特征与传统地图一样，但其形式和处理方式有其自身的特点：

• GIS 中的地图以地理数据库为后盾，可以不受地图分幅的限制，避免地图拼接、剪贴、复制的烦琐。在输出时比例尺也可以在一定范围内调整，不受传统地图固定比例尺的限制。

• GIS 中的地图可以提供远远超过传统地图的内容，供用户选择。在进行制图输出时，根据需要可以分要素、分层和分级提供空间数据，避免了传统地图用不用都一揽子表示的作法，特别有利于要素图和专题图的绘制。

• GIS 中的地图是存于介质上的数据，又具有修改更新软件的支持，只要数据来源通畅，更新起来较为方便。传统地图虽有修测和再版的制度，但更新周期过长，不能表示地面上的最新变化。

• GIS 数字环境中地图符号表达的信息与符号形状的关系不如传统地图明显，计算机不易根据符号的形状判断符号所表达的信息。因此，与符号有关的信息应明确地表示出来。

• 在 GIS 数字环境中，符号化的地图图形在显示时受到限制。最显著的是电子显示设备在水平方向与垂直方向有不同的显示效果，这样对地图图形显示的视觉效果产生不良影响。

• GIS 中制图输出的手段多种多样，有屏幕显示输出、打印机或绘图机纸张输出、印刷胶片输出等。由于这些输出设备在坐标系、坐标单位和颜色空间方面都不同，因此很难做到在不同的输出方式下地图符号显示的视觉效果都令人满意。

2. 屏幕显示输出

其主要功能是在屏幕上显示地图图形，实现空间信息的可视化，它以地图学、色彩学为基础，以可读性为准则，实时、动态和交互性是其主要特征，缩放、漫游和有选择地显示部分内容是其主要技术手段，地图的屏幕显示主要解决以下四个问题：

(1)要素图形关系处理。其目的是处理好图面的一致性和可辨

别性的关系，信息表达和地图感受之间的关系。着重从以下四个方面处理：

①优先级处理。地图要素从其图形表现看有一定的优先级，优先级高的压盖优先级低的，显示过程中优先级的确定可依以下原则进行：

——优先级低的面状要素先画，线状要素次之，点状要素最后画。对面处理时，先理顺面的包含关系，以父面、子面、孙子面的次序对面要素重新进行组织。

——次要的信息先画，重要的信息后画。

——对同一要素层处理时，根据目标的属性确定显示级别，低级别的先画，高级别的后画。如交通层处理时，显示的顺序为：小路、大车路、普通公路、高速公路、最后为铁路。

②显式隐式关系处理。传统地图制图中，由大比例尺地图制作小比例尺地图都要进行制图综合，否则，图面要素太多会影响用户的读图效果。同样道理，在屏幕显示全幅地图时也要按一定的原则去掉一些次要的要素，当开窗放大时，详细的信息才能显示出来。根据地图比例尺、要素类型、要素的质量特征确定要素的优先级，根据显示地图的缩放比例，采用自动计算阈值和人工选择阈值相结合的办法，有选择地显示重要的制图内容。

③目标间关系处理。包括面与面目标，线与面目标，点与面目标，线与线目标，点与线目标之间的关系处理。在过去 DOS 操作系统下，目标间的关系处理通过算法实现，不但处理复杂、费时，而且对原始数据的质量要求很高。目前基于 Windows 的图形显示对于目标间的关系处理，只需利用图形间的压盖关系加上少量的算法，就能达到理想的效果。目标符号化时，被压盖的符号先画，压盖其上的符号用透明色绘制。

④要素特征转换。由于显示的比例尺是灵活可变的，所以要素选择和制图对象的几何特征转换是必不可少的手段。当图形缩小时，原来的线状要素(如半依比例尺桥梁)，按实际尺寸已经无法输

出，必须改用点状符号输出。面状要素（如居民地）当其面积小于一定值时，变成圈形符号表示。

（2）实时响应。实时响应是GIS中电子地图的基本表现形式，灵活的交互功能是其最基本的特征，所以要求屏幕显示能在较短的时间内（通常应不多于2秒）响应用户的干预，使用户没有等待的感觉。为了解决这个问题，数据装入时，要根据目标的坐标数据建立定位索引，再根据一定的算法，利用编程语言提供的图形显示功能，达到快速显示的目的。在地图图形显示时，先确定屏幕上图形变化区域对应的数据区域范围，并根据定位索引读取能覆盖该区域的最小数据集，从而达到灵活、快速、连续的显示效果。

（3）显示区坐标系的建立。计算机的屏幕是有分辨率的，如800×600、1024×768，因此屏幕显示器的可视范围是有限的。而图形符号化又是在一定比例尺空间内进行的，为了使要显示的地图图形能够被全面显示和操作，必须把地图从地图空间变换到计算机屏幕的可视空间，这个变换称为视见变换。

设屏幕可视空间上$M_1 \times M_2$范围内地图的坐标范围是$(X_{min}, Y_{min}, X_{max}, Y_{max})$，则地图上任意一点$(W_x, W_y)$，变为屏幕坐标点$(V_x, V_y)$的转换关系为：

$$
\begin{aligned}
V_x &= \left[\frac{W_x - X_{min}}{X_{max} - X_{min}} \times M_1\right] \\
V_y &= \left[\frac{W_y - Y_{min}}{Y_{max} - Y_{min}} \times M_2\right]
\end{aligned}
\qquad (9\text{-}3)
$$

在上式中假定屏幕坐标系的方向与地图坐标系的方向一致。这时采用屏幕绘图语句和命令，将所有针对(W_x, W_y)的操作都改变为对(V_x, V_y)进行操作，地图图形即可在屏幕上显示。

当屏幕上显示的地图内容不能满足用户的视见要求时，可采用开窗方法放大显示所需部位。这时窗口所对应地图空间范围形成新的$(X_{min}, Y_{min}, X_{max}, Y_{max})$，同样采用上式进行视见变换。同样当对地图在屏幕上进行缩小、漫游显示时，处理的方法与开窗放大

类同。

(4)符号尺寸的确定。为了保持地图显示在屏幕上的清晰性、可读性,其视觉效果必然受到屏幕尺寸和分辨率的影响,屏幕越大、分辨率越高则视觉效果越好,反之则差。由于屏幕尺寸一般都比纸质地图的图幅尺寸小,屏幕上线划也比不上纸质印刷地图的精细程度;又由于人眼看物体时对物体的尺寸是有一定限制的,低于一定限度,符号就不能辨认。这个最低限度约为眼睛1′视角对应的尺寸,因此屏幕上地图符号的尺寸不能设计得太小,要比图式规定的符号尺寸大些,地图图面上内容不能太多,否则就影响地图的可读性。

3. 基于PostScript语言和RIP技术的硬拷贝输出

目前广泛应用的输出设备包括打印机、绘图仪、激光记录仪、直接制版及数字式直接印刷设备等。PostScript语言和RIP技术为更高层次上综合处理文字和图形图像,实现图文合一提供了标准。

(1)PostScript语言和RIP技术概述

①PostScript语言。PostScript语言是美国Adobe公司于1985年推出的具有很强的图形图像处理能力的解释性语言(PostScript Level I),它既是一种通用的程序设计语言,又是一种国际标准的页面描述语言,能对输出复杂页面进行描述,具有图形处理能力强、与设备无关性、易于理解等特点。PostScript语言的最大优点就是能够综合处理文字和图像,在一幅地图当中,可能包含文字、线条、图像、彩色图像、平网等各种要素,PostScript将这些信息形式都用一种计算机数据来表示和描述,为图文合一的版面处理提供了可能。PostScript语言的另一大特点就是避免了版面设计制作对使用设备的依赖性,只要符合PostScript这个标准的数据格式,不管是用Macintosh编辑还是用Sun工作站编辑,都可以通过打印机驱动程序获得共同的PostScript编码文件。PostScript对应的输出设备(打印机、激光照排机等)都可以按照

自身的分辨率进行正确的输出。

1990年Adobe公司在PostScript Level I的基础上推出了PostScript Level II,是目前比较通用的页面描述语言,它对页面的描述是一种高层次的与设备无关的描述。PostScript Level II虽然使用较广,但仍存在许多不完善的地方,对其使用和解释器的研究存在许多问题,如速度慢、易出错等,使用不便。这些缺点和不足主要有下面几方面的原因:

——资料缺乏。目前市场上关于PostScript语言的资料很少,极大地限制了对PostScript语言的应用和开发。

——"方言性"。PostScript语言具有非常明显的方言性,对同一图形图像的描述,可以有多种不同的结果,使软件开发很难做到高效率。

——后缀表示法结构。PostScript语言是解释执行的,其语法采用后缀表示法,即一条语句中,操作数放在相应操作符的前面,使一般操作者对程序的阅读和理解非常困难。

——不具有页面的独立性,不可以对多页面中的某一页进行单独处理。EPS(Encapsulated PostScript)是经封装的PostScript的简称,是所有支持PS技术设备都能识别的文件格式。EPS是PostScript的一种变体,是为支持图形的交换而设计的。它与常规的PS文件的区别是引入了特定的解释字段,排除一些运算符,以及避免一些对大文件有影响的动作。EPS文件格式是由头、PostScript码及尾等构成,它的头必须包含一些DSC(Document Structure Convention,文档结构约定)注释。

Adobe公司利用PS技术的应用软件和输出设备先后进入市场,通过这项技术一般用户在普通的计算机上也可以得到专业用户所需的产品质量,为此PS技术迅速普及开来,成为电子出版界的标准。目前国内有几百套的激光照排机支持PS技术,因此GIS中地图输出的数据使用PS和EPS格式可以最大限度地使用这些输出设备。因此,必须将地图符号化的矢量数据文件转换成PS或

EPS 格式，这样才能实现在 GIS 中的地图输出电子化。

目前国内外大量的出版系统，都是基于 PostScript 语言来设计的。从结构上看，所有系统的基本框架如图 9—29 所示。

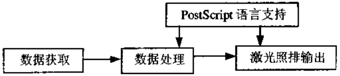


图 9—29 基于 Post Script 语言的出版系统框架

②RIP 技术。PostScript 语言的相关技术包括 RIP 技术、挂网技术、色彩再现、汉字处理等，而 RIP 技术是对 PostScript 语言所描述的对象进行解释的更加复杂的过程。

RIP 英文名称为 Raster Image Processing，即栅格图像处理器，指的是出版系统中可以将用 PostScript 描述的数据解释为点阵图像的硬件设备或解释软件。它在彩色电子出版系统中有非常重要的作用，它的作用是把从工作站或服务器来的页面描述语言快速地解释为可控制激光记录仪输出用的点阵命令，能把 PS 或 EPS 文件中的文字、图形、图像等自动转换成点阵数据，再将点阵数据送到激光照排输出设备上，控制输出设备打点，从而将结果输出到胶片上。

RIP 技术是出版系统中的核心技术，出版系统的开放性、输出质量和输出速度很大程度都取决于 RIP 的优劣。RIP 中的加网处理算法、高速的文字还原技术(包括汉字处理等)是衡量一个 RIP 性能的主要依据。

RIP 一般分硬件 RIP、软件 RIP 和硬件软件相结合的 RIP 。硬件 RIP 的优势是它的速度和稳定性，可靠性也较高；不足之处是成本高，维修困难。软件 RIP 以其灵活性和低价格等诸多优点成为当今世界的主流，它包括软件主体和各种设备的驱动程序两大部分。软件 RIP 具有参数设置灵活、适应性强、控制手段多、升级费用少、安装调试方便等优点，目前国内较流行的 RIP 都是纯

软件 RIP 。硬件 RIP 还原图像跟不上技术的高速发展，随着越来越广泛使用的彩色电子出版系统的问世，硬件 RIP 被软件 RIP 取代已成为必然趋势。

判断一个 RIP 优劣的原则主要有以下几下方面：

——输入、输出插件功能的强弱。好的 RIP 应该具有丰富的输入、输出插件功能，即能接受并解释多种格式的数据，不仅仅只接受最基本的 PS 和 EPS 数据，还应包括 TIFF、TXT 等；输出插件可以提供对一组标准接口的定义，可以驱动符合该标准接口的任何设备。

——输出图像再现的精细程度，特别是色彩传递中色彩的准确程度，保证地图印刷色彩的质量。这在实践生产中非常重要，是衡量 RIP 优劣的主要指标。

——处理速度的快慢。在解释输出时，RIP 的时间长短非常重要。

——支持网络打印功能，这是由出版系统的系统构成所决定的，出版系统离不开网络，必须在网络环境下开发，因为在出版系统中数据的大量交换非常频繁，而且鉴于 Apple 系列的 Macintosh 操作系统和 PC 或工作站中的 Windows 系统共存，所以 RIP 本身必须支持网络打印功能，才能方便应用。

——页面点阵的导入、导出功能以及自动发排和运行条件是否方便灵活等。

(2)基于 PostScript 语言和 RIP 技术进行地图输出的原理

①基本原理。基于 PostScript 语言和 RIP 技术进行地图输出的基本原理是：在 GIS 地图设计与编辑、符号化、注记配置等处理的基础上，实现基于 PostScript 语言的地图数据输出。将满足出版要求的地图数据首先用 PostScript 语言描述，经预打样检查之后，经过地图出版的专用 RIP 处理器，控制输出设备输出专色或分色胶片。地图输出部分的数据流程如图 9—30 所示：

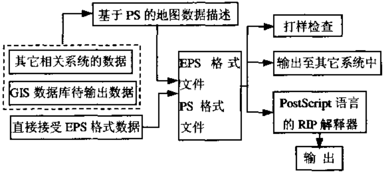


图 9—30 地图输出的数据流程

②工艺流程。地图输出的工艺流程描述如下：

——数据准备。主要是数据库数据基于 PostScript 语言的转换，或者是现有的系统如 ARC/INFO、MAPINFO、MAPGIS 等产生的地图数据向输出控制系统的转换。在这个过程中，经常会出现许多问题，如数据要素丢失，套色不准，色层不统一，符号不全，字体定位偏移等。

——地图编辑。编辑包括地图编辑和地图出版编辑两大部分。地图编辑主要包括比例尺变换、投影变换、地图的拼幅与截幅、图廓尺寸的误差纠正与精度测试、要素的增删改、地图要素综合、图廓整饰等；出版编辑主要包括要素分版、优先级处理、符号配置与生成、符号化处理、色彩设计、汉字注记、地图内容的相互关系处理、区域边界生成、晕渲等。

——色彩模式转换。由于印刷一般要求 CMYK 四色套印，当然地图有其特殊性，有时要求按照不同的要素色依专色来处理，而原始数据，特别是一些图像数据，其色彩模式大多为 RGB 模式，因为 RGB 模式具有较好的色域空间和更广泛的用途，所以必须在输出前完成 RGB—CMYK 的转换，寻求一个高效、准确的色空间转换模型是需要解决的一个主要问题。

——拼版组版等。对于符合出版要求的数据在输出前需要进

行拼版加工，以最大限度地利用输出设备的幅面来输出胶片，在此过程中还有设置印刷套版线、裁切线等一系列处理。

——数字彩色打样。在胶片输出前还要通过数字彩色打样，检查产品的内容、挂网、设色等是否正确，整体效果是否达到用户要求，对存在的问题在进行修改。打样还可以在需要地图份数不多的情况下，输出若干样图直接使用，无需出胶片和印刷。

——RIP 处理。输出控制中的关键部分，主要是解释 PostScript 描述的地图数据，并将数据转换成栅格形式。RIP 过程中要用到许多相关的复杂技术，如挂网，汉字技术等。

——输出胶片。在校样之后，经 RIP 处理并将栅格数据经激光照排输出设备输出四色胶片或专色胶片，供制版和印刷。

为了保证地图输出系统的实用性、先进性以及和其他输出系统的兼容性，要求经编辑处理的地图数据统一存为 PS 或 EPS 格式，而输出控制部分则统一使用 PostScript RIP，以将前端数据解释为可输出的点阵信息，并由此控制照排机等输出设备的动作。基于 PostScript 的 RIP 处理的主要流程如图 9—31 所示。

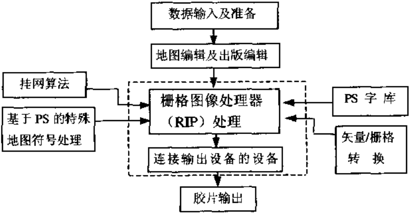


图 9—31 基于 PostScript 的 RIP 处理流程

③关键技术

基于 PostScript 语言和 RIP 技术进行地图输出的关键技术

包括：

- 应用 PostScript 语言
- RIP 处理器设计
- 色彩学及在输出中的应用
- 汉字技术及应用
- 加网技术及应用

——PostScript 语言的使用。PostScript 语言作为一种通用程序设计语言，兼有编程和页面描述功能，但更为重要的是它作为页面描述语言的功能，它可以用来描述图形和文字等，有大量的与图形、图像、文字有关的算符。使用 PostScript 语言首先要提到解释器(Interpreter)的概念，PostScript 解释器是按照用户提供的 PostScript 程序中的指令来控制输出设备的动作，解释器执行页面描述并在打印机、显示器或栅格输出设备上产生输出。

一个结构完整的基于 PostScript 语言的页面描述一般包括两部分：Prolog 部分和 Script 部分。

- Prolog 部分

包括执行 Script 部分时可能用到的过程定义，它是用户生成的 PS 文件的第一部分，也包含与 PostScript 语言支持的输出功能相符的定义。

- Script 部分

由应用程序自动产生，描述生成页的特定要素，它由 Prolog 中定义的过程、PS 操作符、数据等组成。

——RIP 处理器设计。RIP 处理器中包含许多技术，这些措施和技术的使用，最根本的目的是加速和提高地图输出的质量，基本结构图如图 9—32 所示。

——色彩的应用。在地图输出中，有两种颜色呈色机理：1)色光三原色 R(红)、G(绿)、B(蓝)加色法混合呈色机理；2)色料三原色 C(青)、M(品)、Y(黄)和 K(黑)减色法混合呈色机理，这是彩色印刷呈色的基本原理。在输出中经常会出现的问题是：1)RGB 中

实现的颜色在 YMCK 中无法实现；2）YMC 中能得到的颜色，在 RGBK 中无法实现。特别是前者，直接影响输出产品颜色的真实性。

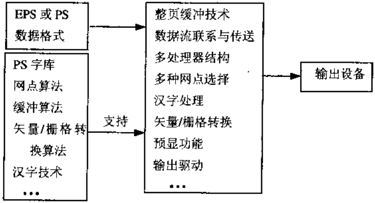


图 9—32 RIP 处理器结构图

可见在地图输出中，一方面必须保证 RGB 色空间到 YMC 色空间的转换尽量做到准确；另一方面，对于地图设计中的设色也提出了新的要求，在设色中要考虑屏幕呈色机理与油墨呈色机理的区别，在颜色设计中应尽量选择色域匹配区域中的颜色，从而可以降低软件对设备的依赖性。

——汉字处理技术。汉字技术主要包括对汉字的表示、引用、处理，特别是在 RIP 中对汉字的处理方法。实际上汉字处理是在 RIP 处理中将汉字曲线描述信息下载转换为点阵信息的过程。

对汉字的应用和处理不同于一般地图线划，很容易出错。一般出版系统对汉字的处理有两种途径：1）在转化为 PS 格式文件时，将汉字的代码和属性表示信息依照汉字曲线描述字库信息转化为相应的曲线轮廓描述（或称为路径），在 RIP 处理过程中汉字等同于一般线划，不再具有汉字的属性。这种方法使数据量增大，效率较低，而且有时会出现字形不美观、锯齿等现象。2）转化为 PS 格式文件时只给出相应字体名称、汉字代码、字变形等属性。在 RIP

处理时才将汉字下载，利用相应的 PS 字库的曲线轮廓描述信息替换，转为点阵信息。这种方法比较科学，处理速度快，输出字的精度高。

——加网技术。挂网是实现地图彩色复制的重要技术手段，是输出控制中 RIP 计算的主要内容。传统工艺中，主要采用接触网屏进行加网拷贝或利用玻璃网屏照相加网的方法获得网点化图像。而在出版系统中，RIP 处理中的挂网主要采用数字挂网技术。

数字挂网一改传统的接触式网屏方式，采用一定的算法对原数据实现栅格化，具体的实现手段是：设定一个加网密度阈值矩阵，实际上是一组数据，该矩阵中存储按照一定的规则从小到大排列（对应阶调的明暗变化）的密度阈值数据，在加网过程中，原数据与该加网密度阈值矩阵相比较，从而对原数据中的值进行比较变化，当数据大于加网密度阈值矩阵时，对应数据为“1”，照排输出时形成一黑点，否则对应数据为“0”，形成白点。

在实际应用中，加网密度阈值矩阵往往选 16×16、8×8 或其他的点阵构成。该矩阵的设计非常重要，对于不同网角、不同形状和大小的网点要有相适应的加网密度阈值矩阵。为了提高效率，增加输出质量，对于该基本实现原理还有很多的优化和改进方案，如多栅格结构的使用等。

④地图输出的特殊性。与其他电子出版物相比，地图有它的独自特点，如自成一体的符号系统、一些特殊的色彩和图形表示方法等。因此，地图的电子出版有其特殊性：

——精度问题。精度问题是地图输出中的一项非常重要的指标，这是由其用途来决定的。高精度要求地图从数据获取、编绘、处理、输出、印刷等各个环节严格控制，显然出版是对其精度影响较大的一个方面。

——地图专用 PostScript 符号库建立。将一些常用的地图符号按 PS 格式矢量化并编码之后存入 PS 符号库，并挂接于 RIP 中。这样，在 GIS 的编辑设计中只需按代码调用符号即可，而将此

符号的查询、数据格式转换及输出交由 RIP 完成,这种方案数据量小,处理速度快,并能保证地图符号的优美。此外,为保证地图注记的质量,还应选用高质量的 PS 字库挂接于专用 RIP 中。

——专用 RIP 的设计。地图出版专用 RIP,不仅能满足普通彩色出版的要求,同时又结合地图的特点添加了一些特殊功能,利用这些功能可以简化地图图形编辑过程、提高地图数据的处理速度、改善地图出版的成图质量。

关于地图出版专用 RIP 的设计思想,可概括为以下几点:

地图出版专用 RIP 应是一个基于 PostScript 语言的解释器。这对于保证地图出版系统的开放性和先进性是很重要的。唯有如此,才能实现地图出版系统与国内外其他各电子出版系统的兼容,便于数据的交换;也唯有如此,才能使地图出版系统随着 PostScript 及相关技术的迅速发展而不断发展,以便紧跟世界先进技术,保证系统的先进性。

引入先进的网点处理(挂网)技术。挂网技术及挂网效果的好坏,直接影响到地图彩色表现和有关层次再现的效果,因此,在地图出版专用 RIP 的研制中应采用国际、国内关于网点处理的一些成熟的和最新的研究成果,如调频网点技术、调幅网中的超细胞结构有理挂网和无理挂网技术等。

增加专用的网线生成功能。国内外现有的 RIP 一般都只有网点挂网算法,而不具备网线挂网功能。而在地图的编制(尤其是专题地图)中,利用网线表示专题要素是经常用到的;地质图和一些专题图的编制中,还保留着利用网线表示色层(甚至叠印生成中间色调)的方法。如果 RIP 不具有网线生成功能,则只有在编图中采用软件功能和手工操作结合的办法画出这些线条,数据量大,处理速度慢,操作也不方便。在地图专用 RIP 中增加了网线生成的功能后,只要在 GIS 的地图编辑中给定有关网线的参数(如线粗、加网线数、角度等)即可,至于如何实现和生成这样的网线则交由 RIP 完成。

——地图出版中的专色处理。地图，特别是地形图考虑到精度问题，在印刷过程中不是采用 CMYK 四色套合印刷获得，而是采用专色印刷，如棕色表示等高线。为了实现专色输出，首先需要将不同颜色的地图数据分离，写入到单独的文件中，如棕色的数据作为一个单独的 EPS 文件或 PS 文件来存储，当然这必须与地图的分层结合起来，否则难以实现单独一个颜色的分离。在 RIP 发排时，RIP 控制输出的胶片不是 CMYK 四色胶片，而是 CMYK 四色按照校对稿的形式输出在一张胶片上，印刷时换上相应的油墨如棕色，从而实现了棕色专色的印刷。有多少个专色，就对应有多少张胶片输出。

(3)PostScript 3 语言和 PDF 文件

Adobe 公司在 PostScript Level I 和 PostScript Level II 的基础上推出了 PostScript 3 语言。PostScript 3 增加了许多新的属性和特点，它支持内置 RIP 补漏白，通过采用与页面无关的 Adobe 的可移植式文件格式 PDF 来解决 PostScript 流程的低效率。PostScript 3 的主要新功能就是能够直接处理 PDF 文件。

PostScript 3 的推出使 RIP 技术大大向前迈了一步，PostScript 3 语言的页面无关性等新特性的应用可以解决过去 RIP 的低效率，使 RIP 的处理更高效、有预期性、更流畅。同时 PostScript 3 的应用也使 RIP 的数据流程更完善，解决了某些现存流程中的问题。

PostScript 3 的最大优点是可以使用多重 RIP 快速地同时处理和打印同一文件中的不同页面。正因为 PostScript 3 的结构是按照高速文件处理的要求来设计的，所以它可以对所接受的 PostScript 文件或 PDF 文件进行并行的 RIP 处理，以便利用多 RIP 同时驱动同一输出设备来达到提高图像处理速度的目的。目前，基于 PostScript 3 的 RIP 产品已经推出。

PDF(Portable Document Format)称为可移植式文件格式，它是一种计算机语言，不是编程语言。PDF 文件的获取是通过由

组好版面的应用程序使用 PS 驱动程序生成 PS 文件，不是送往打印机或绘图仪输出，而是以文件形式保存下来，然后由 Acrobat Distiller 或 PDF Writer 等软件将 PS 文件变成 PDF 文件。过去采用 PS 文件格式存在的最大问题是语言的不可预期性，用户并不知道一个输出过程需要多少时间，以及输出过程是否流畅和稳定，有时打样的字体和色彩与照排输出的结果完全不同。这是因为不同的应用软件采用不同的方式生成 PS 文件，其中有些软件并不理想。另外 PS 文件必须从头到尾进入输出模块，印前的每一次处理都必须读入全部文件并理解其中的内容之后才可进行处理，修改非常繁复。

PDF 文件较 PS 文件更可靠，预期性更强，文件格式始终如一。PDF 文件的优点非常突出：

——PDF 作为计算机语言是高度构造化的，不能编程组装，所以构造完全相同的文件用不同的应用软件来制作，再将其结果转换成 PDF 格式时，所得到的 PDF 文件几乎是一样的，没有 PS 语言的“方言性”。

——PDF 与 PS 语言不同，它具有页面独立性，可以独立处理页面。PDF 能够以页为单位记述各个页的内容，各页面所共有的项目能够分别独立保存，可以独立地对多页 PDF 文件中的某一页进行单独处理，无需处理整个文件。

因此，支持 PDF 文件格式的 PostScript 3 与前两个 PostScript 版本相比，具有更大的优点。PostScript 3 和 PDF 经过集成后可创造出一种供传输和出版的数字化文件，并成为硬拷贝数字化输出的标准。

§9.7 多种形式的空间信息可视化

可视化(Visualization)在信息世界中具有特殊的地位。在人机交互中，视觉是信息传输和接收的主要渠道，尤其对于多维信息

可视化具有独特优点，而空间信息正是多维的。空间信息可视化是指运用地图学、计算机图形学和图像处理技术，将地学信息输入、处理、查询、分析的数据及结果用图形符号、图像，并结合图表、文字、表格、录像等可视化形式显示并进行交互处理的理论、方法和技术。采用声音及触觉、嗅觉、味觉等多媒体方式可以使空间信息的传递、接收更为形象、具体和逼真，从广义上理解，可听、可嗅、可味、可触也可以纳入可视化范畴。

1. 空间信息可视化的形式

地图是空间信息可视化的最主要的形式，也是最原始的形式。在计算机上将空间信息用图形和文本表示，在计算机图形学出现的同时就出现了。这是空间信息可视化的较为简单而常用的形式。多媒体技术、地图动画技术、三维技术和虚拟现实技术的产生和发展，使空间信息可视化进入一个崭新的时期。可视化的形式主要有：

——地图：包括纸质地图和电子地图。这两种形式仅是计算机上数字地图的硬拷贝和软拷贝的差别，硬拷贝的如纸质地图，软拷贝的如电子地图比前者具有更多的优点：制作灵活，形式极其多样，修改方便，周期短，色彩丰富，动态性强，查询方便。

——多媒体信息：使用文本、表格、声音、图像、录像、音频、视频等各种形式，综合、形象地表现空间信息，是空间信息可视化的重要形式。

——动态地图：动态地图的产生和发展是时空 GIS 发展的必要基础和前提。它是利用地图动画技术，直观而又逼真地显示地理实体运动变化的规律和特点。

——三维仿真地图：三维仿真地图是基于三维仿真和计算机三维真实图形技术而产生的三维地图。

——虚拟现实：虚拟现实是空间信息可视化进一步研究和发展的新形式。它是由计算机和其他设备如头盔、数据手套等组成的高级人机交互系统，是以视觉为主，也结合听、触、嗅甚至味觉来感

知的环境，使人们有如进人真实的地理空间环境之中并可与之交互作用。

2. **多媒体信息**

近年来，多媒体技术得到迅速发展，多媒体的应用更以极强的渗透力进入入类工作与生活的各个领域，多媒体时代的来临，为人们勾画出一个多姿多彩的视听世界。多媒体技术，简单地说就是由计算机综合处理文字、图形、图像、动画、声音及视频等信息，使其在不同的界面上流通，特别是指具有传输、转换及同步化的功能，也就是说由计算机同时抓取、操作、编辑、存储或显示不同媒体形态的能力，而达成人机交互的操作环境。

(1)多媒体数据描述与表示方法。从表达形式来看，空间信息的多媒体形式主要包括文本、图形、图像、视频、声音等五种媒体及其组合，它们既是空间信息可视化的主要内容，也是实现 GIS 中数字地图可视化的基本技术和方法。借助于上述五种媒体技术，可以使用户直观、形象、生动地阅读和使用数字地图。我们主要讨论这五种媒体形式在空间信息可视化中的描述与表达方法。

①文本媒体及其表达。文本是人们最为熟悉，也最为常用的一种空间信息表示方法，如各种文字介绍、区域环境解说、系统标题、功能菜单、补充说明、地理名称、现象或事物的质量或数量特征标识等都是信息系统的重要内容，文本媒体是构成地理信息系统信息库的主要内容，如地名数据库、统计数据库等。

文本媒体的组织结构可按如下方式定义：

```
typedef struct Text{
        Char    Text—Header;     /* 文本标题
        Char    Text—Version;    /* 版本号
        Char    Text—Date;       /* 创作日期
        Char    Text—Author;     /* 著作者
        Char    Text—Content;    /* 文本内容
        Char    Text—Type;       /* 文本类型
```

```
        Long     Text—Pointer;     /* 文本指针
    }
```

其中,文本标题主要表明文本的名称,如"北京市概况"等。文本内容可以是西文、中文,其中还可嵌入图形、图表、声音等其他媒体。文本指针主要是为定义热字链接而设计的。

②图形媒体及其表达。图形实际上是用矢量数据表示和存贮的地图图形,有时也称其为矢量图形。地图图形实际上是空间点集在一个二维平面上的投影,它们可以分解为点、线、面三种基本图形元素,每一种图形元素又包括几何位置及其属性特征。在计算机中可用如下方式进行描述和组织:

点:用一对 X,Y 坐标表示,即:(X,Y);

线:用一串有序的 X,Y 坐标对表示,即:$\{(X_1,Y_1),(X_2,Y_2),\cdots\cdots,(Xn,Yn)\}$;

面:用一串有序的、而且首尾坐标相同的 X,Y 坐标对表示其轮廓范围,即:

$$\{(X_1,Y_1),(X_2,Y_2),\cdots\cdots,(Xn,Yn),(X_1,Y_1)\}。$$

③图像媒体及其表达。图像是表征空间环境现象的一种重要媒体形式,多媒体系统中的图像主要包括卫片、航片、照片和各种栅格图像等。在多媒体系统中所使用的图像通常都用"位图"(Bitmap)来存贮和生成。位图是一个用来描述构成图像的所有像元及其特征的信息矩阵。

图像媒体可用如下方式给予描述:

```
typedef Image {
    Char         image_header;         /* 图像标题
    Short        Length, Width;        /* 图像长度和宽度
    Int          Depth;                /* 图像深度
    PIC_FRAM     Data[1024][768];      /* 图像大小
  }
```

④视频媒体及其表达。视频是一种非常重要的动态视觉媒体,

它能够以生动、真实的形式再现某一区域或某一现象的实地状况，从而增强了空间信息表示的动态感和真实感。视频的获取及处理也是所有媒体中最复杂的一类。

视频媒体可按如下方式描述。

```
typedef Video{
        Int     Sample—freq;        /*采样频率
        Int     Quantitive;         /*采样质量
        Char    Video—Type;         /*视频类型
        Char    Video—Header;       /*视频标题
        Char    Video—Intro;        /*视频内容介绍
    }
```

⑤音频媒体及其表达。声音是多媒体系统中不可缺少的内容和组成部分。声音主要包括视频的背景音乐和文字介绍录音等两种，它们都属于数字音频媒体，是空间信息多媒体数据处理的重要内容。

数字音频处理的主要内容包括：声音预处理和模数转换，其基本过程是：音频采集设备将接收到的声音信号送至隔离放大器进行放大，再经衰减器和低通滤波器处理后送模数转换器处理，获得数字音频信号，最后经并/串行转换变为串行数据后送用户使用。

多媒体系统中所使用的声音文件主要包括两种，即波形文件(Wave)和音乐设备数字接口(MIDI)。其中使用最多的是 Wave 文件。Wave 文件是资源交换文件 RIFF(Resource Interchange File Format)的一种类型，RIFF 文件是以块为单位进行组织的层次结构——块中包含块，被包含的块中又包含块。Wave 文件最少包含三个块，RIFF 块是其中最大的，整个 Wave 文件就是一个 RIFF 块。第二和第三个块称为子块，包含在 RIFF 块之中，它们是 fmt 子块和 data 子块。fmt 子块保存 Wave 文件的采样描述信息，包括格式标识符、声道数、每秒采样数、平均每秒采样字节数、块对齐标志和采样位数等信息，这些信息决定了对 data 子块中数据的

解释方式。data 子块是文件的最大部分，紧跟在 fmt 子块之后，包含所有数字波形数据。

（2）多媒体数据获取。文本的输入可以采用扫描输入的方法，用文字识别软件实现扫描后图像文字的自动识别；可以采用语音输入的方法；也可以用文字编辑软件进行录入和编辑。

图形的输入可通过图形工具 Coreldraw 等软件制作或通过图形商品库得到。

图像的输入方式可以有：彩色扫描已有照片；直接获得数字影像；从商品化的素材库中选取合适素材。

视频的输入可以把来源于摄像机、录像机和视频光盘的模拟视频信号经过计算机处理转换为数字视频信号，也可以利用专业视频制作软件由静态图像合成得到。

音频的输入可以通过音频卡录制声音；购买商品化的音效库；从已有的数字音频、CD 中截取或转换。

（3）多媒体数据压缩。多媒体数据的数据量非常之大，如果不进行数据的压缩，势必影响到使用的速度和效率。多媒体数据的压缩主要涉及图像、音频和视频的压缩。

图像的数据量主要与图像大小和存储格式有关。要减少图像的数据量通常是使图像尺寸缩小，图像存为 JPG 格式比存为 TIFF 格式数据量少近 10 倍，虽然对数据有所损失，但图像质量也能得到保证。

对于数字化音频 Wave 文件，其数据量主要由采样频率、采样位数和声道数决定。采样频率越高、采样位数越大、取两个声道采样，则数据量越大，反之则越小。对于解说词声音可以采用较低的采样频率、8 位采样位数和单声道，以减少数据量；而对音乐声音则采样频率高一些为好。对于音频的压缩，也可以考虑使用音乐设备数字接口（MIDI），MIDI 其实不产生任何声音，它只是计算机、合成器、键盘和其他音乐设备之间相互通信的一个协议，因此是最经济的多媒体声音类型。例如，保存一小时一个合成器的音调仅需

要6个字节，相比而言，一小时最低保真度的 Wave 数据需要 39，690，000 字节。

压缩视频文件数据量的方法通常是设定压缩格式和改变窗口大小等。

(4)多媒体数据管理。空间信息的可视化采用多媒体表示的关键问题之一就是对项目中众多的媒体数据进行管理，这包括对文本、图形、图像、音频、视频等文件的管理。对它的管理有两种方法：一种是使用数据库管理系统(DBMS)；另一种是建立特殊的逻辑目录结构。目前多采用第一种方法进行多媒体数据的管理。

为方便多媒体数据库管理，对各种媒体形式采用了统一的数据库结构，分别把文本、图形、图像、音频、视频文件名作为索引，对应于相关的媒体文件。数据库结构简洁明了，利用 DBMS 的库管理功能可方便地录入、增加与维护，为数据的增加、删除、修改奠定了良好的基础。

(5)多媒体数据集成。多媒体数据集成是利用程序设计语言或现成的商业软件，调用需要的媒体数据和设计用户界面，完成空间信息可视化的多媒体系统。多媒体数据集成的方式有：

——复式多媒体集成。一般用于制作拥有一系列类似媒体文件的多媒体系统，常见的集成软件有 TookBook，HyperCard。

——图标式多媒体集成。这种方式是按对象的顺序，以流程图为主干完成多媒体系统，常见的集成软件有 Authorware，IconAuthor，方正奥思等。

——时间为基础的多媒体集成。这种方式是以时间顺序来组织媒体数据，完成多媒体系统，常见的集成软件有 Director，Action 等。

——程序设计语言多媒体集成。这种方式是以高级的程序设计语言为基础，开发面向对象的程序，从而生成多媒体系统。这种方式由于灵活性大，适于开发多种类型的多媒体系统，尤其适于开发大型的较为复杂的多媒体系统，常用 Visual C++，Visual Basic，

Delphi 等程序设计语言。

3. **动态地图**

动态地图的产生和发展是时空 GIS 发展的必要基础和前提，是空间信息可视化中一个蓬勃发展的分支，空间信息时空变化的抽象、夸大、取舍、化简等与动态地图的设计和制作直接相关。

(1)动态地图的特征和作用。动态地图的主要特征是逼真而又形象地表现出空间信息时空变化的状态、特点和过程，即运动中的特点。具体而言，动态地图可以用于：

——动态模拟：使重要事物变迁过程再现，如地壳演变、冰川的形成、人口增长与变化等。在这些复杂的动态过程中，动态地图是一个有力的表现方式，它可以通过增加或减低变化速度，暂停变化以仔细观察某一时间断面，改变观察地点和视角，获得运动过程中的各种信息。

——运动模拟：对于运动的地理实体，如人、车、船、机、星、弹等，其运动状态的测定和调整，以及环境的测定和调整，可以由动态地图来帮助完成。

——实时跟踪：在运动的物体上安装全球定位系统，能显示运动物体的运动轨迹，使空中管制、交通监控和疏导、战役和战术的合围等具有可靠的时空信息保证。

(2)动态地图的表示方法。为了表示空间地理实体的运动状态和特点，可采用各种方法及组合。

——利用传统的地图符号和颜色等表示方法。采用传统的视觉变量组成动态符号，结合定位图表、分区统计图表法以及动线法来表示。这方面在军事上应用较多，行军、战斗突击、战役迂回等动态过程都需要采用传统的动线法予以表示。

——采用定义了动态视觉变量的动态符号来表示。基于动态视觉变量：利用视觉变量的变化时长、速率、次序及节奏等，可设计相应的动态符号，并加上电子地图的常用手段：闪烁、跳跃、色度和亮度变化等，来反映运动中物体的质量、数量、空间和时间变化特

征。

——采用连续快照方法制作多幅或一组地图。这是采用一系列状态连续的地图来表现空间信息时空变化的方法。这一方法在状态表现方面是比较全面的，但对变化表达不够明确，同时数据冗余量大。

—— 地图动画。其制作方法与上一方法相似，仅仅是它适当地在空间差异中内插了足够多的快照，使状态差异由突变改为渐变。这一方法弥补了上一方法变化表达不够明确、时间维上拓扑关系模糊的缺点，是动态地图表现较为丰富的形式，缺点是数据量大。

4. 三维仿真地图

空间信息的可视化在早期受限于计算机二维图形显示技术的发展，大量的研究放在图形显示的算法上，如画线、颜色设计、符号填充、图形打印等。继二维可视化研究后，进一步发展为对地学等值面（如数字地面模型）的三维图形显示技术的研究，它是通过三维到二维的坐标转换、隐藏线、面消除、阴影处理、光照模型等技术，把三维空间数据投影显示在二维平面上。由于对空间数据场的表达是二维的，而不是真三维实体空间关系的描述，因此属于 2.5 维可视化。但现实世界是真三维空间，二维可视化无法表达诸如地质体、矿山、海洋、大气等地学真三维数据场，所以三维 GIS 及空间信息可视化的形式——三维仿真地图成为当前 GIS 的研究热点。随着计算机硬软件技术和科学计算可视化技术的飞速提高，空间信息的可视化由二维向三维发展成为可能。

(1)空间信息的三维表达。目前，GIS 和计算机辅助设计中，对真三维空间数据通常采用面模型(Surface Graphics)的方式，即由矢量边线构成物体，通过显示深度表达，用消隐、浓淡技术表达若干面之间的前后关系。任何对地理场景、视角、光线的变化，都需要重新计算。

与面模型相对应的是体模型(Volume Graphics)，即应用三维立体格网来直接绘制地理场景，空间中的每一个格网对应某一个

可量测的属性值。三维立体格网单位称为体素(Voxel),体素可以看作二维像素在三维空间的推广。把采集到的空间数据矢量值分配给体素,其值通过体素的材质、颜色、纹理或透明度反映出来。

由于基于体素的图形可以直接对数据进行操作,同时具有实时动画的功能,在运算上具有较大的优点,它已成为科学可视化研究和开发的重要方向。

(2)三维仿真地图的制图设计。与传统二维地图的设计不同,三维环境下的制图设计受立体感受的影响,除了采用传统的视觉变量表达空间目标的特征外,还要用色彩、阴影、纹理、透视变换等来构造三维影像,这些变量被称为图形深度要素。

色彩和阴影用来模拟光照效果,因为虽然三维视觉影像效果是由光线照射到物体上留下阴影形成的,但计算机屏幕模拟的本质是由色彩的饱和度和亮度变化形成的,物体本身的色彩、反射光色彩、照射光色彩的不同可形成截然不同的效果。纹理则是在三维影像表面贴上与实物表面相似或相同的图像,以提高三维仿真地图的真实感。透视变换则是通过对空间目标的坐标变换来模拟不同的观察角度,例如俯视、侧视等。采用图形深度要素进行地图设计是三维地图设计中最重要和最基本的问题,因为三维地图的视觉效果主要依靠这些要素来表达。

(3)三维仿真地图的显示技术。由于计算机硬软件技术的飞速发展,可用于三维地图显示的三维图形工业标准相继推出,提高了三维图形开发效率。为了提高三维图形显示速度,可以从两个方面考虑:一是压缩数据量以缩短数据的访问时间;二是提高实时显示速度并模拟"越近越清晰"的视觉规律。三维地图数据的数据量大最主要是由于数字地形的引入,对于数字地形,尤其是 DEM 数据,可以采用线性四叉树和二维游程编码相结合来进行压缩。

过去在数字地形显示时,当要进行局部观察时,常采用机械放大方法,以致于图像越来越粗糙,为了解决这一问题,可以采用 LOD(Level of Detail)技术,即细节层次模型方法。LOD 技术的基

本思想是：根据不同的显示要求采用不同精细程度的模型来表示同一物体，物体细节越丰富，模型越复杂；物体细节越粗略，模型越简单，从而可以根据显示要求选择相应精细程度的模型，大大降低场景的复杂度，使实时绘制场景成为可能。

数字地形在显示时除了采用 LOD 技术来提高视觉效果外，还采用“雾化”技术来增强效果。“雾化”是指在计算机三维地图仿真中模拟烟雾、薄雾、浓烟和污染等大气效果，使离视点较远的物体开始淡化成雾的颜色。可以通过判断视点与地形参考点间的距离，采用不同分辨率来显示数字地形，距离越远，分辨率越低。为了弥补数字地形分辨率变化造成的失真，可以采用“雾化”的技术来屏蔽，既可以大幅度提高数字地形的显示速度，又提高了地形观察模拟的真实感。

(4)三维仿真地图的产品类型。对于空间地形信息的三维表达，由于地形数据是在三维形式下给定光源由计算机自动模拟，效果逼真并且易于调整，可以产生不同类型的三维地图产品。这些产品类型有：

——将正射投影的 DEM 进行分层设色并取消高程值的比例(将所有高程值设为 0)，则可形成分层设色地势图。

——对正射投影的 DEM 采用光照模型并调整光照方向，则可产生晕渲图；

——当调整视点位置时，对 DEM 进行投影变换并适当夸大高程比例并进行光照和色彩设计，可以得到三维透视立体图。

——在数字地形上叠加矢量要素或其他专题要素时，不仅可以获得矢量专题要素随地形变化的直观起伏，而且可获得三维专题图。

——把航片或遥感影像叠加在数字地形上，可获得真实感的地景模型。

5. 虚拟现实

虚拟现实(Virtual Reality，简称 VR)又称临境技术或人工环

境，是指通过头盔式的三维立体显示器、数据手套、三维鼠标、数据衣、立体声耳机等使人能完全沉浸于计算机生成创造的一种特殊三维图形环境，并且人可以操作控制三维图形环境，实现特殊的目的。虚拟现实是发展到一定水平的计算机技术与思维科学相结合的产物，它为人类认识世界开辟了一条新途径，为空间信息的可视化提供了一种新形式。

（1）虚拟现实的基本特征。交互（Interaction）、沉浸（Immersion）、想象（Imagination）通常被认为是虚拟现实的三个基本特征。

①交互。虚拟现实的最大特点就是用户可以用自然方式与虚拟环境进行交互操作，这种人机交互比平时计算机屏幕界面交互要复杂得多。例如，当人在虚拟场景中行走时，体位和视角的任何变化，引起场景画面的变动，计算机要连续不断地重新构造画面。如果场景中的物体又是移动的（如坦克、飞机），其动态三维构图的计算量非常大。

②沉浸。虚拟现实的沉浸特征可以看作是交互的深化，即置身于一个"适人化"的多维信息空间，以人在自然空间所具有的各种感觉功能（视、听、触、体、嗅）去感知虚拟空间的信息。在这个空间中，技术的难点是感知系统、肌肉系统与VR系统的交互，只有各种感觉的逼真感受，才能产生沉浸于多维信息空间的仿真感觉。

③想象。虚拟环境的设计不仅来自于真实世界，即仿制客观世界现有的物体、现象、行为等，而且可以来自人的想象世界。这个想象世界可以是将难以在现实生活中出现的微观、剧变、艰险、复杂的环境，用虚拟现实技术再现出来，使用户达到亲历的机会。

（2）虚拟现实系统的基本构成。虚拟现实系统是一个软、硬件与多种传感器结合的高科技系统，由于在该系统中尚存有大量未解决的难题以及理论的和技术的盲区，其构制的难度是显而易见的。一个基本的VR系统应该由VR生成系统、软件及开发工具、知觉系统三部分组成。

①VR 生成系统。VR 生成系统包括：

——图像生成器。这是构成虚拟现实的核心内容，因为人接受外界信息的 80%是通过视觉获得的。它应具备将图像按双眼视差理论分解为双像的功能，以供使用者产生空间立体感。

——音频系统。VR 系统中音响效果是与图像配合，由音频器提供的，其效果取决于各种动态音响的建模质量。

——跟踪及定位。用户体位的变动包括头部的活动或手的动作的位置信息必须立即反馈给计算机，以便实时改变图像和音响。最常用的方法是提供一个可以包含用户在内的三维空间传感器，即一个空间坐标系统，以获得用户相应部位的 x，y，z 坐标和围绕三坐标轴的旋转角。

②用户的知觉系统。也就是人机交互的各种界面。

——声像头盔。头盔是将双眼立体观察屏幕、立体声耳机与定位接收器组合在一起的硬件装置，是目前系统中能产生最佳沉浸效果的组合。

——大屏幕显示器。在大屏幕上播放可分为左右眼分别接收的重叠图像，用户用立体眼镜来观看，由一个红外发射器控制双眼的液晶开关，产生立体视觉。

——数据手套。是最常用的人机接口工具，可以把手的动作反馈到计算机以体现与图像视体的接触、抚摸，它是一个装满传感器和利用各种微量量测装置的高技术集成产品。

——跟踪球。这是一种控制相对位置和方向的传感器，通过它用户可以改变虚拟图像中物体的位置和方向。跟踪球是单向控制器，无法接收虚拟图像变化的反馈，通常配合其他接口设备使用。

③软件及数据支撑。一个虚拟现实系统应有图像软件环境支撑。通常可分为两类：开发工具包和创作系统。虚拟现实软件的开发需要具备实时系统、面向对象的语言、网络、造型等多方面知识，面对于急于应用 VR 技术于各专业领域的人员，最好的办法是使用 VR 开发工具，包括三维建模软件、实时仿真软件以及相应的函

数库等。创作系统是带有图形用户界面的完整软件，通过它只需简单编程即可生成虚拟场景。

在虚拟地形环境开发中，地理空间数据库的支持特别重要。地理数据库是空间信息基础设施之一，是大投资、长时间建立的工程项目，是一种信息资源。地理数据库以地形数据为主，包括地形、水系、居民地、交通线、地物的三维数据等，是生成空间定位地形图像的基础。与之相配合的是地面影像数据库，这是根据已定位的航空照片与卫星照片数字化而成的，是构成地形三维图像的重要数据来源。

§9.8 GIS中地图制图的常用算法

1. 点值图的计算机制作

点值图是一种特殊的统计型专题地图，它的制作方法与普通统计型专题地图的制作方法有所不同，因此，在这里作一介绍。

点值图是用一定大小、形状相同的点来表示现象的分布范围、数量特征和分布密度的方法，即是一种面状要素数值表示的符号化方法。点值法分为均匀布点法和定位布点法，这里介绍均匀布点法的计算机制图方法，其基本步骤为：

确定总点数→确定点值→确定每个区域的点数→对每个区域均匀布点

点值法表示的是某种面状要素的精确数据值，如各地区的人口数据。也就是说，布点是在一定范围内进行的，这个范围可以是行政区划单元或其他地理单元。通常为了尽可能使所布的点与实际情况相一致，总是采用较低一级的区域范围。所以，用计算机绘制点值图时必须已知每个区域的范围（即多边形坐标值）和相应的指标的数值（如人口数）。

设制图区域范围的面积为 P(km^2)，比例尺分母为 M，则制图区域范围的面积 P′为：

$$P' = 10^{12} \times P \times M^2 (mm^2) \qquad (9-4)$$

若点与点之间的间隔为 d'，点的直径为 d(通常取 0.4—0.6mm)，则在 P'内应放的点数 N 为：

$$N = P' / (d + d') \qquad (9-5)$$

d'通常取 0.2 毫米，即每点的四周各有 0.1 毫米的空白。知道了点数 N 后，若专题内容指标的总值为 A，则点值 S 可用下式计算：

$$S = A/N \qquad (9-6)$$

使用时应对 S 取整。设某个地区的指标值为 A_0，则在该地区应布的点数 N_0 为：

$$N_0 = A_0/S \qquad (9-7)$$

知道了某个区域应布点数 N_0 后，就可调出该多边形的数据进行均匀布点了。均匀布点的基本方法是，用间隔 D(大于点的直径 d)的水平方向的平行线分割该区域，根据平行线与区域的交点，计算出区域内线段的总长度 L。因为要在这些平行线上配置点，则平行线上每两点距离 D_0 为：

$$D_0 = L/N_0 \qquad (9-8)$$

这样在平行线上每隔距离 D_0 配置一个点，直到把 N_0 个点配置完，就完成了布点。但这种方法由于是事先确定了平行线的间隔，并不能真正保证均匀布点，只是在水平方向上和垂直方向上各保持均匀。为了解决这个问题，可以设想用边长为 D 的正方形来填充该区域，若每个区域内放一个点，则可以做到均匀分布，这时正方形的个数就是点的个数 N_0，而正方形的面积之和应等于该区域的面积 A_0，则正方形的边长 D 为：

$$D = \sqrt{A_0/N_0} \qquad (9-9)$$

这样以 D 为水平方向平行线的间隔来分割该区域，就比用指定间隔的平行线来分割更能体现出均匀分布的要求。

当不知道制图区域的面积时，可以根据构成制图区域的多边

形来求出其面积，设多边形的平面坐标为$\{(x_i,y_i),i=1,2,\cdots,n\}$，则该多边形的面积A可用如下公式求出：

$$A=\frac{1}{2}\sum_{i=1}^{n}((y_{i+1}+y_i)\cdot(x_{i+1}-x_i)) \qquad (9-10)$$

其中当$i=n$时，$x_{n+1}=x_1$，$y_{n+1}=y_1$；$i=1$时，$y_0=y_n$。

2. 等值线的追踪

等值线可以用于表示三维空间或可以认为是三维空间的现象，如地形、气压等。计算机绘制等值线分为三角网DEM(TIN)追踪等值线和网格型DEM追踪等值线，分别介绍如下。

(1)三角网DEM(TIN)追踪等值线。

——数据准备。在离散点构成了三角网之后，就存在点位高程数据和三角网数据。点位高程数据：

$$\begin{array}{lll} x(1), & y(1), & z(1) \\ x(2), & y(2), & z(2) \\ & \vdots & \\ x(N), & y(N), & z(N) \end{array}$$

其中，x，y，z是地形点的三维坐标，可以分别读入三个数组，N为地形点总数。

三角网数据：

$$\begin{array}{lll} t_1(1), & t_2(1), & t_3(1) \\ t_1(2), & t_2(2), & t_3(2) \\ & \vdots & \\ t_1(L), & t_2(L), & t_3(L) \end{array}$$

其中，t_1，t_2，t_3为三角网中每个三角形的顶点，L为生成的三角形的总数。

对上述数据进行处理后，可以找出与每个三角形相邻的另外三个三角形的编号，记录在给定的数组中。如果某一三角形只与一个或二个三角形相邻(这种情况只出现在三角网的边缘上)，则相应三角形的相邻三角形的编号记为0。某个三角形的相邻三角形

用 h_1,h_2,h_3 数组记录，形成相邻三角形数据。

相邻三角形数据：

$$
\begin{array}{ccc}
h_1(1), & h_2(1), & h_3(1) \\
h_1(2), & h_2(2), & h_3(2) \\
 & \vdots & \\
h_1(L), & h_2(L), & h_3(L)
\end{array}
$$

其顺序号即为三角形的编号。

——等值点的确定

一个三角形的边上是否有等值点通常包含下列几种情况：

①若三角形的三个顶点的高程相等，则三角形的边上无等值点。如果三顶点的高程等于等值线的高程，则三顶点就是等值点，由于顶点可能被三个以上三角形共用，所以不考虑这种情况。如图9—33①。

②若三角形三顶点的高程值不相等，则当每条边两端点高程满足：

$(Z-Z_1)(Z-Z_2)\geqslant 0$ 时，则该边无等值点，否则必有等值点。

$(Z-Z_1)(Z-Z_3)\geqslant 0$ 时，则该边无等值点，否则必有等值点。

$(Z-Z_2)(Z-Z_3)\geqslant 0$ 时，则该边无等值点，否则必有等值点。

其中，Z_1,Z_2,Z_3 为三角形三顶点的高程值，Z 为等值线高程。

通常一个三角形如果有等值点，则有两个，如图 9—33②所示。

③若三角形三顶点高程不等，而其中有一个顶点的高程等于等值线高程，则如果该三角形还存在另一个等值点，必然位于该顶点的对边上。如图 9—33③。

④若三角形有两个顶点高程相等，则该三角形如果存在等值点，必位于靠近第三点的两边上(如图 9—33④)，或者该相等的两顶点就是等值点(如图 9—33⑤)，后一种情况不考虑。

在确定了三角形上存在等值点后，可用内插法求得等值点的坐标，其插值公式可写为：

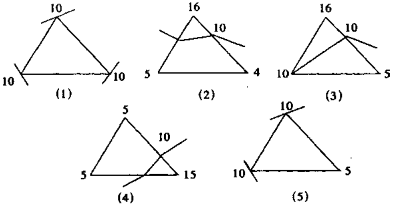


图 9—33　确定等值点的几种情况

$$X=X_1+\frac{X_2-X_1}{Z_2-Z_1}(Z-Z_1)$$

$$Y=Y_1+\frac{Y_2-Y_1}{Z_2-Z_1}(Z-Z_1) \qquad (9—11)$$

其中(X_1,Y_1),(X_2,Y_2),(X_3,Y_3)为三角形的三顶点坐标,相应高程为Z_1,Z_2,Z_3,等值线的高程为Z,(X,Y)为所求的等值点。三角形其它边上的等值点坐标也可用类似公式求出。

——绘制等值线的步骤

①对于一给定的高程值,在三角网中先找到该等高线通过的第一个三角形。一般情况下可在该三角形边界上求出两个等值点。该三角形编号为T。

②以T三角形中的一个等值点为起点,沿外方向追踪等值点,即根据向外方向相邻三角形的编号调用该三角形相应边的数据,算出另一个等值点的位置;再由该等值点所邻三角形求出下一等值点,直到等值线闭合(最后追踪回到T号三角形)或为开曲线的端点(三角网边缘相邻三角形编号为0)为止。

③对于开曲线,以T号三角形的另一点为起点,沿另一方向

追踪等值点，重复上述工作，直到该曲线绘制完。

——注意事项

①三角形三顶点编号与相邻三角形编号对应，即 t_1 点对应 h_1 号三角形，t_2 点对应 h_2 号三角形，t_3 点对应 h_3 号三角形，如图 9—34 所示。

②追踪等值点时，应根据出发点所对应的三角形顶点的点位选择相邻三角形作为追踪三角形。记正在追踪的三角形为 K。

③当 K＝T 或 0 时，追踪结束。但对于 K＝0 的开曲线，还应转向另一方向追踪，两方向追踪完毕才结束。

④当出发点为三角形某一顶点时，取与之不等高的顶点所连边线的相邻三角形为追踪三角形。

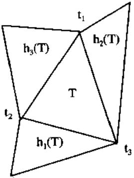


图 9—34　三角形顶点与相邻三角形对应关系

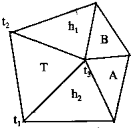


图 9—35　三角形顶点为等值点的追踪方法

⑤在图 9—35 中，设追踪的三角形为 T，顶点为（t_1，t_2，t_3），其中 t_3 为等值点。根据④可知向 h_1 或 h_2 追踪，设向 h_2 追踪，但 h_2 没有等值点，则追踪方向以 t_3 为轴心，依次向前推进，即顺序取 A、B 为追踪三角形，直到找到等值点为止，因为该等值点一定存在。

（2）格网 DEM 追踪等值线。格网 DEM 追踪等值线的方法早在六七十年代就已出现，方法有多种多样。下面介绍一种简捷、效果较好的方法，即方向关联法。

——有关概念。

对于格网 DEM 数据，每一个数据点与周围 8 个点相邻（格网边的点除外）。设格网 DEM 是 m×n 的数据阵列，沿 X 方向的分割记为 j=1,2,…,n；沿 Y 方向的分割记为 i=1,2,…,m。则某一点(i,j)与周围点的关系可由图 9—36 所示。

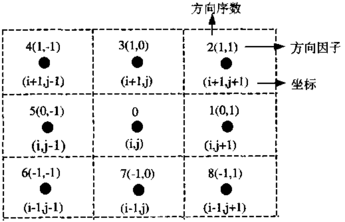


图 9—36　格网 DEM 数据点与周围点

方向序数——方向序数表明了点(i,j)与周围点之间的方向关系，从右邻点开始，按逆时针方向分定别定义为 1,2,…,8。

方向因子——方向因子用于描述点(i,j)与周围点之间的坐标关系。对应于方向序数 1,2,3,4,5,6,7,8，其方向因子为：(0,1),(1,1),(1,0),(1,−1),(0,−1),(−1,−1),(−1,0),(−1,1)。已知某点坐标为(i,j)，则该点的相邻点的坐标可由该相邻点的方向序数对应的方向因子加上原始点的坐标(i,j)即可算出。例如某点的坐标为(18,15)，则方向序数为 4 的邻点的坐标为(19,14)。

设追踪的等值线的高程为 Z，则若格网点的高程小于或等于 Z 则称为内点，否则称为外点，等值线必然在一对相邻的内外点之间通过。

——追踪等值线的方法。在地图学中，起始并终止于区域边界的等值线称为开曲线，在区域内闭合的等值线称为闭曲线。

①等值线起始点的搜索。开曲线需要沿区域的四周边界搜索起始点。即在满足 $i=1$ 或 $i=m$ 或 $j=1$ 或 $j=n$ 时，相邻边界点构成一个内外点对。

闭曲线应在区域内部搜索起始点。即在满足 $1<i<m, 1<j<n$ 时，相邻点构成一个内外点对。

②终止条件。开曲线的终止条件是被插值的内外点对同时为边界点。

闭曲线的终止条件是，被插值的内外点对与起始内外点对相同。

③内插点坐标的计算。设内插点的内点位置为(i,j)；其外点的方向序数对应的方向因子为(a,b)；等值线的高程为 Z；格网间距为 d；内点高程 Z_1，外点高程为 Z_2。则内插点的坐标(x,y)可按下式计算：

$$\begin{cases} x=d\cdot(j+b\cdot(1-\dfrac{Z-Z_2}{Z_1-Z_2})) \\ y=d\cdot(i+a\cdot(1-\dfrac{Z-Z_2}{Z_1-Z_2})) \end{cases} \qquad (9—12)$$

④追踪等值线的步骤。

· 搜索等值线起始点。

先搜索开曲线的起始点，再搜索闭曲线的起始点。当起始点的搜索遍布了整个区域后，结束。

在找到一个起始点后，应计算内插点的坐标；并对起始点的内点作标记，防止重复追踪；同时记录下内点到外点的方向序数，设为 N。

· 追踪等值线的下一点。

设当前等值点的内点为(i,j)，外点的方向数为 N。

计算新的方向序数，令 N＝N＋1，并把新的方向数 N 归化到

8 以内，即：

$$\begin{cases} N=N, & \text{当 } N\leqslant 8; \\ N=N-8, & N>8; \end{cases}$$

找出新方向序数对应的邻点的高程，判断是内点还是外点。

若是内点：对原内点作标记，把该点作为新内点，并令新内点的方向序数为 N＝N＋5，并同前面一样归化到 8 以内。然后在新方向序数对应的内外点之间计算内插点坐标；若是外点：直接计算出内外之间的内插点坐标。

· 判断这条等值线是否终止。

判断条件如前所述。若未终止，转第二步；若已终止，转第一步。

⑤示例。如图 9—37 所示，现对 4×4 的格网 DEM 追踪等值线，等值线的高程为 5。为了叙述的方便，对应格网点编号为 P_1，P_2，……P_{16}，格网点的高程值在圆圈内标出。

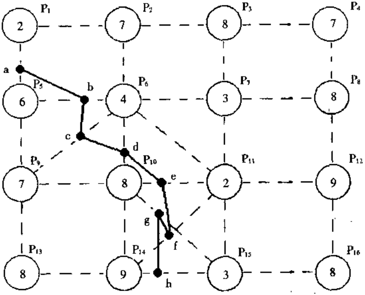


图 9—37 格网 DEM 追踪等值线示例

首先找等值线的起始点。设在P_1P_5间找到一对内外点对，且P_1是内点，则方向序数N＝7，可方便地计算出内插点a的位置。

令：N＝N＋1＝8，对应的邻点为P_6。由于P_6也是内点，因此对P_1作标记，把P_6当作新的内点。同时，令N＝N＋5＝13，归算到8以内为N＝5。对P_6点，方向序数为5的邻点为P_5。在P_6P_5间找出插值点b。

令：N＝N＋1＝6，对应的邻点为P_9。P_9是外点，因此在P_6P_9之间找出插值点c。

令：N＝N＋1＝7，对应的邻点为P_{10}。P_{10}是外点，因此在P_6P_{10}间找出插值点d。

令：N＝N＋1＝8，对应的邻点为P_{11}。P_{11}为内点，因此对P_6作标记，把P_{11}作为新的内点。

因此，令N＝N＋5＝13，归算到8以内为，N＝5。对P_{11}，方向序数为5的邻点为P_{10}。在$P_{11}P_{10}$间找出插值点e。

令：N＝N＋1＝6，对应的邻点为P_{14}。P_{14}为外点，因此，在$P_{11}P_{14}$间找出插值点f。

令：N＝N＋1＝7，对应的邻点为P_{15}。P_{15}为内点，因此，对P_{11}作标记，把P_{15}当作新的内点。同时，令N＝N＋5＝12，归算到8以内为，N＝4。对P_{15}，方向序数对应的邻点为P_{10}，在$P_{15}P_{10}$间找出插值g。

令：N＝N＋1＝5，对应的邻点为P_{14}。P_{14}为外点，因此在$P_{15}P_{14}$间找出内插点h。

最后，由于符合等值线终止条件而完成了该条等值线追踪，然后继续下条等值线的追踪。

第十章　GIS的开发

许多人相信21世纪最重要的产品是——信息。而大量的信息是和地理位置有关的，地理信息和地理信息的处理系统——GIS已经成为一项巨大的产业。但是就在10年前，人们对GIS基本上还是一无所知。国内的GIS产品，像MapGIS、GeoStar，还不存在；在《计算机图形图像学报》上刊登GIS广告更是闻所未闻；每年组织一次GIS年会和测评也显得可笑。但是，短短的几年时间里，这些都已成为现实。GIS好象无所不在，如今，GIS已经成为生活中一种非常现实的技术。在很多情况下，人们把他们的工作、他们的旅行、他们的娱乐、他们的健康、他们的决策都依赖于GIS和地理数据。

GIS软件在这10年间发生了很大的变化。硬件性能的提高，软件技术的演进，计算机体系结构的不断变化，均促进了更为强大、成熟和更为复杂的GIS软件系统的出现。如果一个GIS系统开发成功，那么，这种复杂性能够产生奇迹般的结果，但是它们也给建造这些复杂系统的人员带来很多的问题，给开发造成很大的难度。

本章介绍的方法和技术是想要建造正确的GIS软件产品所需要用到的。这些方法和技术其实就是我们即熟悉又陌生的软件工程，它包括GIS软件开发中采用的方法，应用的技术和工具。

软件工程——将系统化的、规范的、可度量的方法用于软件的开发、运行和维护的过程。(IEE93)

§10.1 项目开发方法

在计算机发展的早期阶段，系统开发的主要预算在于硬件，对于硬件的质量控制和质量保证有非常详尽的工程化方法。软件的开发则相对简单，没有什么系统化的方法。软件环境是一种个人化的环境，很容易设想，一个程序员会在半夜两点起来去完成一个苦思冥想的问题，至于软件设计，仅仅是他头脑中的一种模糊想法，这种想法及其实现当然也不会形成任何规范的文档。这时候的软件开发基本上处于混乱的无管理状态，由于系统的简单和参与共同开发的人员很少，软件的开发也通常会取得成功。

随着计算机系统的增多，计算机软件库开始扩展。这时，程序员发现他们越来越繁忙起来：当发现错误时需要纠正所有这些程序；当用户需求发生变化时需要修改；当硬件环境更新时需要适应。这些活动统称为软件维护。在软件维护上所花费的精力开始以惊人的速度消耗资源，其费用甚至远远超过了系统初始开发的整个费用。

一般说来，GIS 是一个大型、复杂的软件系统。为了确保这样一个大型软件项目的质量，降低开发费用，提高系统开发的成功率，必须借助科学的软件开发方法。有效的系统开发需要开发小组的共同努力，了解软件开发的方法对参与 GIS 开发项目的所有人员都是非常重要的，其中包括用户和所有 GIS 专业人员。

开发方法并不是一开始就有，在软件开发还是一种"艺术"的日子，谁也不会去想一些规则来约束那些随心所欲的"艺术家"，这时候的软件往往都是一些较小的单独的成果，很少需要互相的协调和交流。也许只有第一个需要组织大批人员的开发项目出现以后，才会产生与"艺术"相悖的一些普遍的规范化"方法"。任何共同开发过程中产生的错误和教训都是宝贵的经验，这些经验如果形成文档，并且在开发组织中发布，要求所有成员共同遵守，就会形

成一种事实上的标准。现在,所有成熟的、成功的标准都已经写进了教科书,发达的计算机工业还提供了很多有用的工具,因此,当我们在抱怨 GIS 系统开发的复杂和昂贵时,至少有一点是可以庆幸的——不需要自己去发明开发方法。

以下是一些软件开发通用的开发方法。他们也同样适用于 GIS 的开发。

1. 线形顺序模型

有时也称为传统的生命周期法(Life Cycle),瀑布法等。它是在第二和第三代编程语言(COBOL,FORTRAN 等)时代设计出来的。指的是从开发过程一个阶段的输出流到下一阶段的线性的、顺序的方法。整个软件过程历经系统调查、需求分析、系统设计、编码、测试和维护等阶段。有时,某一阶段得到的信息会使周期向前一阶段回退(如图 10—1)。

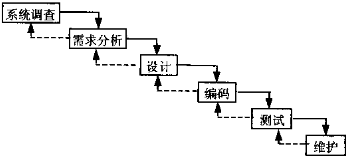


图 10—1 传统的系统开发生命周期法

这种方法存在的最大问题是,用户只有在系统几乎全部完成时才能使用该系统。因此很难以最小的开销得到满足业务需求的产品。如果用户一开始难以清楚地给出所有需求或开发小组对用户需求的理解有偏差,对已经煮成"熟饭"的系统的任何改动都将付出非常大的代价。因为越在系统开发周期的后面阶段发生错误,不但有些前面的部分工作要重做,而且这个错误所影响的人员也

会越多。这种方法的另一个问题是开发者常常因为在生命开发周期的某个环节发生"阻塞"而影响随后环节的进行，很容易理解，如果在系统设计阶段发生延误，那么严重依赖于它的编码阶段将不能进行，事实上，花在等待上的时间可能会超过花在开发上的时间。

为了尽量避免这种方法存在的问题，有经验的系统开发者会在项目生命周期早期就尽可能的去发现错误和风险，"如果你不主动地去处理项目中的风险，他们就会主动地袭击你"(Tom Gilb)。通常，在系统整个开发过程中，会形成大量的文档或中间产品，在每一阶段结束后，通过对它们严格的检查并且经常与用户交流，可以了解它们是否达到需求，从而获得对系统质量的最大控制。

不管怎样，这种方法已经非常成熟和形式化，是软件工程中应用最广泛的过程模型，在软件工程中占有非常重要的位置。对于一些大型的复杂的系统，经常会用到这种方法，特别是一些政府或部队的大型项目。

2. 原型法(Prototyping)

当不知道做什么时做什么呢？也许原型法就是回答。当用户定义了系统的一般性目标，但不能标识出详细的输入、输出、处理及反馈需求，这时可以先建立系统的一个初始版本提供给用户试用，经过用户反馈后，进行改进后成为第二代、第三代版本，一直到整个系统最终完成(如图 10—2)。

创建原型的方法主要有两种。第一种是先完成系统的核心部分，再逐步增加其他的功能组件；另一种方法是先完成所有功能组件的主要部分，再逐步增加次要部分的功能。两种方法都是根据前面的反馈进行原型的调整和精化，整个过程是迭代的。

原型法的优点在于它能够很快完成可操作原型并提供给用户，这样用户会变得更积极主动，容易及时发现错误并判断是否满足需求。

当然，在很快看到原型开始工作运行的"愉悦"过后，用户和开

发人员会发现，为了使原型尽快的工作，最初采用的程序设计语言可能不合适（如使用 Visual Basic 而非系统要求的 Visual C++）；最初采用的算法效率可能太低或精度不能满足要求（采用的原因仅是因为这种算法是现成的）。久而久之，开发者可能会忘记他们最初选择的弊端或者已经习惯于这种选择。更严重的是，系统质量和长期的可维护性可能就根本未予考虑。这样，用户就会不停的要求修改，而开发者会因为并没有为改进留有足够的预算等原因而抵制或限制用户的修改要求，如果不能很好的折衷，这种争论会变成无休止的矛盾。

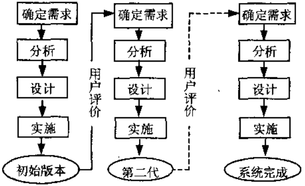


图 10—2 原型法

虽然会出现问题，原型法仍然不失为一种有效的项目开发方法。关键在于用户和开发者必须达成一致：原型被建造仅是为了定义需求，之后就该被抛弃（或部分被抛弃），因此需要澄清原型好坏及成功的标准。

3. 快速应用开发（Rapid Application Develop，RAD）

政府部门的决策效率有时会比预想的要低，而一旦作出决策，他们又希望能尽快地看到成果，尽管预算资金还没有到位；现代商业要适应信息化的环境，要有极强的应变能力，有时要求一个系统在 6 个月内完成。这样的需求要求能非常快地生成一个“功能完善的系统”，快速应用开发（RAD）就是适应这种需求的方法。

为了加速系统创建过程，RAD 项目经常会运用已设计好的技术、方法及工具，有些项目会采用基于组件的建造方法。市场上有很多支持快速应用开发的软件工具（象 Powersoft 的 PowerBuilder、Microsoft 的 Visual Basic 等）。在 RAD 项目中，用户的参与程度远比其他方法要高，在整个开发过程中，用户要花费大量的时间和开发人员一道参加系统的设计和实现。

由于 RAD 强调极短的开发周期，因此，这样的项目一般压力很大，每天的工作时间较长，对开发者的心理和生理都是巨大的考验。如果 RAD 项目要用到很多新技术，那么其技术风险也会很高，失败的可能性也会增大。

4. 迭代增量模型（Iterative and Incremental）

线性顺序模型是假设当线性序列完成之后就能提交完善的系统。但实际上，在软件开发的过程中，需求会经常发生改变，象瀑布一样的直线开发过程往往不实际。原型法虽然通过迭代能很好的帮助用户和开发者理解需求，但本质上，它并不是交付一个最终产品系统。

迭代增量模型融合了线性顺序模型的基本过程和原型法的迭代特征。当使用迭代增量模型时，第一个增量往往是核心的产品，经过用户的使用和评估再进行下一个增量的开发，这样不断的重复直到最终产品的完成。从总的过程来看，它和原型法一样是迭代的，不同的是，迭代增量模型强调每一次迭代都产生一个可操作的产品。从它每一个增量的实现过程来看，又是线性的，都通过分析、设计、编码和调试过程，产生一个可发布的“增量”。（如图 10—3）

迭代增量模型把系统实现的技术风险分解到每一个增量，在早期的增量中可以避免使用不成熟或技术难度大的软硬件。从而使整个系统的技术风险变得最小和可控制，在系统开发的早期，还能缓解资金和人员的紧张。迭代增量模型的另外一个好处是，由于不断产生可运行的增量，这对增强用户和开发人员的自信心和继续开发的兴趣是很有帮助的。

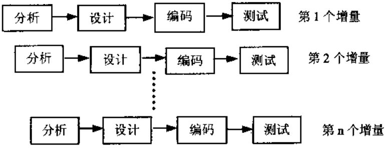


图 10—3 迭代增量模型

5. 面向对象(Object－Oriented)的方法

我们生活在对象的世界里，面向对象的方法是一种模型化世界的抽象方法，它按照人类认识世界的思维方式分析和解决问题，通过引人对象、类、消息等基本概念和机制来灵活的模拟真实世界的状态和行为(面向对象的基本概念见 § 3.3)。

面向对象的方法有很多内在的优点，其中之一是对象重用，这样可以进行快速的软件开发。另外，面向对象的软件易于维护和扩展，因为它的结构是内在松耦合的。

曾经有很多年，"面向对象"指用面向对象的程序设计语言(如 C^{++}，Smalltalk 等)的软件开发方法，今天，面向对象技术已经包含完整的软件工程观点，它包括面向对象分析(OOA)、面向对象设计(OOD)和面向对象计算机辅助软件工程(OOCASE)等完整的软件工程的方法和工具。到了 20 世纪 90 年代，对象技术和面向对象的软件工程已经被很多软件、特别是信息系统软件所采用。

对象技术的流行衍生出许多面向对象的分析方法，比较著名的有 Booch 方法(重视操作和行为，对类和继承的概念也值得借鉴)、Rumbaugh 方法(重视数据和对象的建模)、Jacobson 方法(面向用例)和 Code/Yourdon 方法(一种简单明了的方法)等。这些方法都引入一个项目的系统分析的过程，都有自己面向对象建模的符号体系，即建模语言。由于众多面向对象方法一直独立发展并且都有各自的优缺点，用户很难根据自己的应用特点选择合适的方

法。这些方法谁优谁劣的争论持续了很长时间，争论的焦点集中于建模语言。直到 1995 年，Grady Booch、Jim Rumbaugh 和 Ivar Jacobson 这三个世界著名的面向对象技术专家终于走到一起，并推出了统一建模语言（UML，Unified Modeling Language）。到 1996 年底，UML 已经占领了面向对象技术市场的 85%，成为可视化建模语言事实上的工业标准。1997 年 11 月 17 日，OMG 接纳 UML 作为基于面向对象技术的统一建模语言。1998，UML 的设计者们基于 UML 语言，综合他们各自的系统分析过程，推出了一种命名为"统一软件开发过程"（Unified Software Development Process）的软件开发方法。这是一种通用的方法，适用于各种各样的软件系统、应用领域和开发组织。

6. 统一软件开发过程（USDP，Unified Software Development Process）

统一软件开发过程是一个基于组件和 UML 语言的软件项目开发方法，它的思想可以归纳为三点：用例驱动（Use－Case Driven），以体系结构为中心（Architecture－Centric），迭代和增量（Iterative and Incremental）。

（1）用例驱动（Use－Case Driven）。软件是为用户而存在的，这里用户包括人，也包括与软件交互的其他系统。一个用例就是软件系统与用户完成一次交互的功能。用例从用户的观点对系统的功能建摸。因此用例构成了一个用户和系统开发者都能接受的基本概念。

为一个系统定义的用例是整个开发过程的基础。所有用例组成了用例模型，它是需求分析的结果。用例模型取代传统的功能规范，描述了完整的系统功能。功能规范回答的问题是"系统能做什么?"，用例的策略在于"系统能为每个用户做什么?"，用例驱动整个项目的开发过程（设计、实施、测试），开发人员随后创建设计和实施模型来实现用例，测试人员根据用例建造测试用例，看实施模型的组件是否正确实现了用例。用例就象一条"红线"贯穿于整个

软件开发过程。

虽然用例驱动软件工程，但用例的选择并不是孤立的，它驱动系统的体系结构，系统体系结构则影响用例的选择。它们随着系统开发生命周期的延续而逐渐成熟。

(2)以体系结构为中心(Architecture－Centric)。一个建筑物用写景、立面、电缆、管道、通风和暖气等不同的建筑蓝图来描述建筑物的不同方面，来展现建筑物的体系结构。同样，软件系统的体系结构也可以通过不同的视图来描绘，比如软件系统的功能、软件系统的逻辑构成、软件在系统平台的物理分布等等。对于复杂的软件项目，不使用体系结构，或者使用差的体系结构，将冒很大的技术风险。

体系结构是理解和解释系统如何工作的设计视图，它描述的是系统的主要特性(主要的类，主要的进程、线程，持久化机制，通信机制，分层和子系统，接口)，而不考虑其细节。

参与软件项目的人员很多，以软件的体系结构为中心，不同的人员对软件体系结构的不同方面感兴趣。系统分析员使用它组织和描述需求，理解技术约束和风险；用户可以从高层看到它们定制的是个什么样的系统；软件项目经理用它来组织小组的开发计划；设计人员理解基本的原则并定位他们自己设计的边界；其他开发组织用来理解如何与它交互；设计师用它进行扩展和重用的推理。建筑蓝图已经形成标准，它使得投资商、设计人员、建筑工人和客户在阅读和使用蓝图的时候不会有障碍。为了能够在不同的人员之间进行通信、回顾、评论和改善，也应该使用一种标准的、一致的表达方法来描述软件项目的体系结构。

(3)迭代和增量(Iterative and Incremental)。统一软件开发过程不是在项目结束时一次性提交软件，而是分块或分成一些“迷你”项目逐次开发和提交。整个开发过程是一系列周期的重复循环，每一个周期包括开始(Inception)、加工(Elaboration)、构造(Construction)和移交(Transition)四个阶段，每个阶段由若干次

迭代组成(如图 10—4)。每一次迭代又都要经历分析、设计、实现和测试的工作流程。

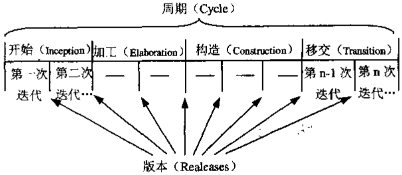


图 10—4 一个周期的阶段和迭代

7. 为 GIS 项目选用开发方法

通常,选用软件项目的开发方法是非常主观的,它受用户需求、个人(组织)喜好、现有的开发工具以及对改变的畏惧等多种因素的影响,当软件开发组织的成员熟悉了一种开发方法以后,简直就象一种信仰,不容易轻易改变。因此,对于整个项目开发组或软件开发组织来说,已经形成标准化的开发方法往往被认为是最好的。但如果开发的项目需要广泛使用新技术(如面向对象技术或组件技术),或对用到的开发工具或数据库管理系统做了重大修改,就要考虑采用新的项目开发方法。

GIS 近年来的趋势之一是由传统的对抽象的点、线、面关系描述向对真实世界的真实目标的描述与管理转变。支持上述变化的技术之一就是面向对象技术及面向对象的空间数据库的日益实用和成熟。随着 UML 语言的出现并且成为面向对象的标准建模语言,采用基于 UML 语言的方法和 CASE 工具进行面向对象的 GIS 软件开发已经成为新的主流。

§10.2 工具型GIS的开发

按照GIS的功能，GIS可分为工具型和应用型两种(见§1.4)。由于工具型GIS和应用型GIS的功能特点不同，一般采用不同的开发方法。在这一节中，我们以统一软件开发过程(USDP，Unified Software Development Process)为例，介绍工具型GIS的开发过程。

1. USDP的模型元素(Modeling Elements)

USDP开发过程包括四种主要的模型元素：工作人员(Workers)、行为(Activities)、成果(Artifacts)(如图10—5所示)、流程(Workflows)。

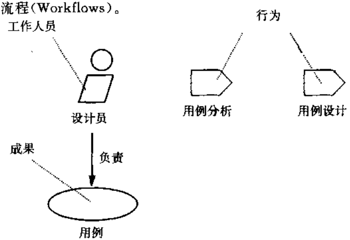


图10—5 工作人员、行为和成果

(1)工作人员(Workers)。GIS一般是个大型的、复杂的软件工程项目，参与项目的人员比较多，如何组织他们实现有效的软件开发是项目成败的关键。有效的系统开发需要所有开发参与者的共同努力，参与软件开发的人员通常可以分为项目管理者(Project Manager)、系统分析员(System analyst)、设计人员(Designer)、程序员(Programmer)、支持人员(Supporter)、资料员

(Documenter)、客户(Customer)和最终用户(End User)如图10—6)。如果投资和系统代价非常高时,开发组织的领导往往也会加入进来,成为项目的一部分。

· 项目主管(Project Manager)

对项目主管而言,项目的管理即是对人的管理,如何招募、培训、报酬以及培养团队精神和企业文化是他们的实践领域。GIS 的项目主管并不一定是 GIS 专家,因为他的工作大多集中于人的活动,因此能胜任 GIS 软件开发的人员并不一定具有对人管理的领导能力。最大限度地发挥每个人的技术、创造力和潜力,很好的组织、计划和控制对项目的开发,这才是项目主管的任务。一个好的项目主管应该使 GIS 开发人员全身心地、愉快地投人到软件开发中,而不应该有后顾之忧。

· 系统分析员(System analyst)

在 GIS 项目中,系统分析员(高级的系统分析员有时称为首席设计师或总师)是精通 GIS 业务分析和设计的专业人员,负责领导和协调制定需求、对系统的定界等。他们在与项目主管、软件开发小组、支持人员、客户的交流中发挥不同的作用。

· 软件开发小组(Software Development Team)、小组负责人(Team Leader)、设计人员(Designer)、程序员(Programmer)

在一个大的软件项目中,软件开发小组是软件开发组织中的最小单元,开发人员被分成若干个小组,每个小组完成一个或多个系统的功能组件或模块。

软件开发小组的组织方式有很多种,规模也可大可小,这取决于项目的管理风格、项目的难易程度、项目的模块化程度以及小组中开发人员的数目等等。有些小组没有固定的负责人,问题的讨论和解决方法的确定由小组成员共同决策,如果能够很好的调动小组成员的主动性,这种很松散的组织方式有利于创新和技术上的突破。但如果是在开发过去已经做过的类似项目时,松散的小组组织则效率低下。这时小组中往往有一个固定的负责人,他与项目主

管一起负责本组与其他小组的通信，并计划、协调小组中每个开发人员的技术活动。

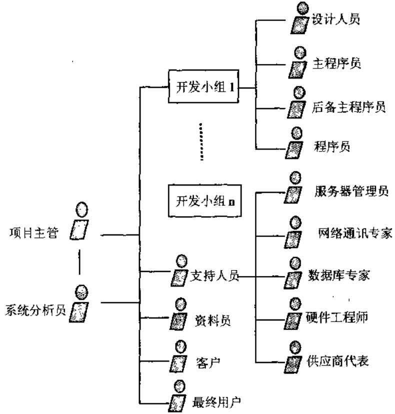


图 10－6　GIS 项目参与者的结构层次图

在 GIS 项目开发小组中，成员通常包括设计人员、主程序员、后备主程序员和其他程序员。设计人员负责定义 GIS 类的功能、操作、属性以及类之间的关系，并且决定如何进行类的修改以适应开发环境。测试设计人员要负责拟订测试计划和测试模型，实施测试过程，评价测试的范围、结果和有效性。主程序员要求具有较强的业务组织能力和丰富的编程经验，能够解决开发中的一些疑难问题，主程序员往往参与整个 GIS 项目的设计和计划，并制定本

小组的开发计划。后备主程序员支持主程序员的活动，并能在项目进行过程中，以最小的代价取代主程序员的工作。其他程序员按照主程序员制定的计划开发和修改软件。

不管 GIS 开发小组采取什么样的组织方式，如果他们有同一的对于成功的定义，有团队精神，有共同的目标和文化，那么其强大的凝聚力会产生巨大的动力，并产生很高的软件生产率。

· 支持人员(Supporter)

GIS 专业领域之外的其他专家，可能包括服务器管理员、网络和通讯专家、数据库专家、硬件工程师以及软件工具供应商的代表等。

· 资料员(Documenter)

为多个小组服务，维护和控制所有的软件配置及其安全，包括文档、源程序、数据等；帮助采集、收集和转换数据；分类和索引可重用软件组件等。

· 客户(Customers)

负责说明待开发 GIS 项目的软件需求的人员。

· 最终用户(End user)

一旦开发的 GIS 项目成为正式的产品，最终用户是直接与软件进行交互的人员。

软件开发小组必须建立有效的方法，以协调项目参与人员之间的关系，必须建立小组成员之间及多个小组之间的通信机制，象个人之间的交流以及定期的会议等。对于制定的工作计划或出现的问题，应该作为打开的状态列入会议文档，并在下一次会议中进行状态复审，看上次打开的状态是否已经关闭。在项目结束的时候，所有打开的状态都应该关闭。

有时候也会召开不定期的会议去寻找解决问题的办法，但又不能占用太多的时间。从这一点考虑，应该只找相关的人参加相关的会议，或者向有关人员分发会议的文档，开发人员能够从文档中获得他们感兴趣的信息。

全球信息高速公路的出现以前所未有的方式开放了全球的工作，象 GIS 这样大型的项目往往涉及不同地方的人员参与，Internet 给地理分布的开发人员提供了召集会议的有效手段(电子邮件、电子公告栏、视频会议等)。有些疑难问题甚至可以得到项目之外的专家支持，在 Internet 上，软件工程师可以参加与他当前关心的技术领域相关的新闻组，在新闻组中张贴的问题可得到全球"好事者"的回答。有调查表明，在所有协调和通信中，与项目之外的人进行讨论最有价值。

(2)行为(Activities)。在面向对象的术语中，工作人员是一个活动的对象，工作人员的行为是指对象执行的操作，以下是一些行为的例子。

- 计划一次迭代：由项目主管执行。
- 发现用例及其执行者：由系统分析员执行。
- 设计审阅：由设计审阅员执行。
- 性能测试：有性能测试员执行。

行为可以分步骤进行，主要有以下三类步骤：

- 思索(Thinking steps)

  理解任务的性质，收集和检验输入的成果，设计输出。
- 执行(Performing steps)

  创建并更新成果。
- 审阅(Reviewing steps)

  对照标准对结果进行检查。

以发现 GIS 中地图注记用例及其执行者这一行为为例，其步骤如下，其中 a、b 和 c 步骤需要思索，d 和 e 是执行步骤，步骤 f 则要求对行为的结果进行审阅。

① 发现地图注记的执行者(如 GIS 最终用户)。

② 发现地图注记用例(如对点要素进行注记)。

③ 描述地图注记执行者和用例的交互(GIS 最终用户如何进行点要素注记)。

④ 对地图注记用例及其执行者进行打包(将地图注记的执行者和用例分成一组)。

⑤ 用用例图表示用例模型(画出地图注记的用例图)。

⑥ 对地图注记的用例模型进行评价。

(3) 成果(Artifacts)。成果是指软件开发过程中产生、修改和使用的信息。是软件项目中有形的产品。它可以是一种行为的输入,也可以是行为的结果和输出。在面向对象的设计术语中,成果是行为的参数,它有各种表现形式:

· 模型,例如用例模型或设计模型。

· 模型元素,例如类,用例或子系统。

· 文档,例如软件结构文档。

· 源代码。

· 可执行文件。

当成果是项目的外部输入或者是一些简单的描述信息时,成果可能是纯粹的文档。但大多数情况下,USDP 并不鼓励把什么都打印到纸上,相反,用创建成果的工具对其进行维护反而是最有效、最实际的管理方法,并且在交付时能确保信息的现势性和准确性,例如储存在 Rational Rose 的设计模型,储存在 Microsoft Project 中的项目计划,储存在 Requisite Pro 中项目需求等。当然,在需要的时候,用这些工具能以快照的形式自动生成文档。

USDP 把成果分成五类,或者说是信息集,这些成果随着迭代开发周期不断地发展:

· 管理集,与项目管理相关的成果,例如软件开发计划、状态评价等。

· 需求集,与软件系统定义相关的成果,例如业务模型、用例模型。

· 设计集,包含对系统的描述,例如设计模型,系统框架描述,测试模型等。

· 实施集,包括源代码、可执行文件等。

· 部署集,包括安装脚本,用户手册以及培训材料等。

(4)工作流程(Workflows)。单单只有工作人员,行为和成果并不能构成完整的开发过程,我们还需要一系列的行为来描述工作人员之间的交互关系,来产生我们需要的成果。我们把这种行为系列称为工作流程。USDP 核心的软件工程工作流程主要有:

· 业务建模
· 需求分析
· 分析和设计
· 实施
· 测试
· 部署

在 UML 术语中,工作流程可以表示成顺序图表、合作图表或者行为图表。以下我们将以行为图表为例。

2. GIS 开发的工作流程(Workflow)

(1)GIS 业务建模(Business Modeling)。在开始 GIS 系统开发以前,通常要进行业务建模,产生的业务用例模型和业务对象模型能保证客户、最终用户和开发人员对 GIS 业务领域有相同的理解,并且可以从中获得系统的需求。

——工作人员。在 GIS 的业务建模中,包括以下工作人员:

· GIS 业务过程分析员(GIS Business-process Analyst)。负责对 GIS 业务领域进行简略的描述和定界,领导和协调 GIS 业务用例建模。定义 GIS 业务用例、执行者及其交互。
· GIS 业务设计员(GIS Business Designer)。详细描述 GIS 业务用例,指定业务工作人员和业务实体,定义业务工作人员的责任、操作、属性以及相互之间的关系。确定需要提交的成果。
· GIS 业务模型审阅员(GIS Business Model Reviewer)。审阅业务建模的成果。对业务用例、业务执行者、业务实体的描述进行审阅。

——成果。GIS 业务建模包括以下成果：

· GIS 业务用例模型(Business use—case Model)。GIS 打算实现的功能模型。如图 10—7 所示。

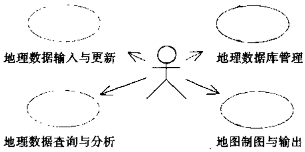


图 10—7 GIS 业务用例

· GIS 业务对象模型(Business object Model)。描述实现 GIS 业务用例的对象模型。

——工作流程。GIS 业务建模的工作流程如图 10—8 所示。

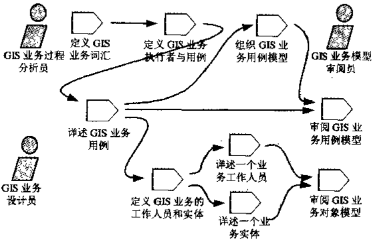


图 10—8 GIS 业务建模工作流程

(2)GIS 需求分析(Requirement)。需求分析是软件工程过程的第一个步骤，目的是确定用户、风险承担者的需要，GIS 要解决的问题等，据此确定系统的轮廓和目标，还包括 GIS 要实现的功能及处理过程、性能需求、设计约束和校验标准等。需求分析的调查方法有很多，最常见的是直接询问用户和风险承担者，但如果用户和风险承担者对需求不明确，对 GIS 的功能不了解，就需要系统分析员去引导、挖掘他们的要求，也可以采取座谈、填表、与用户一起工作等方法，使系统需求的描述最接近于用户的想法。

需求分析的第一步是定义用例，以描述 GIS 系统的外部功能需求。用例模型的主要元素是用例、执行者和系统边界。用例用于描述每个功能需求，系统边界用于界定系统的功能范围，而执行者用于描述执行系统外部功能的有关实体。绝大多数用例可以在系统需求分析阶段确定，但随着系统的进展，可能会发现更多的用例，甚至会发现前面定义的用例存在不够确切或错误的地方，需要重新修改。因此，在整个系统开发过程中，都应该时刻关注用例。

——工作人员。在 GIS 的需求分析中，包括以下工作人员：

· GIS 系统分析员(GIS System analyst)。负责对 GIS 的功能进行简略的描述和定界，领导和协调 GIS 用例建模并导出需求。

· GIS 用例说明员(GIS Use—case specifier)。对 GIS 的系统功能、用例的需求等方面进行详细描述。

· GIS 设计师(GIS Architect)。负责定义重要的用例和需求。

· 用户界面设计员(User—interface Designer)。负责定义、设计用户界面并建立用户界面的原型。

——成果。GIS 需求分析包括以下成果：

· GIS 用例模型(Use—case model)。由 GIS 业务用例模型进一步导出，工具型 GIS 的用例图可能如图 10—9 所示。

· 补充规范(Supplementary specifications)，包括一些不适合用用例描述的详细需求。

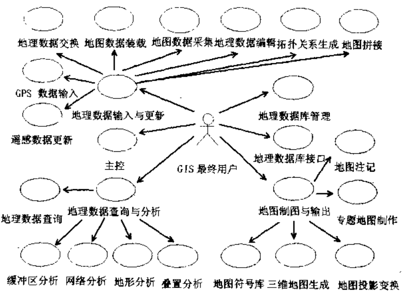


图 10—9 GIS 用例模型

· GIS 需求(Requirements attributes)。

· GIS 用户界面原型(User—interface prototype).

——工作流程。GIS 需求分析的工作流程如图 10—10 所示。

(3)GIS 分析与设计(Analysis and Design)。分析和设计阶段的任务是通过综合考虑所有的技术限制,阅读和分析需求说明,以扩展和细化需求分析阶段的模型。设计的目的是指明一种易转化成代码的工作方案,是对分析工作的细化,即进一步细化分析阶段所提取的类,并且增加新类以处理诸如数据库、用户接口、通信、设备等技术领域的问题。

——工作人员。在 GIS 的分析与设计中,包括以下工作人员:

· GIS 设计师(GIS Architect)。和其他工作人员相比,设计师对 GIS 业务应该有很宽但并不一定很深的知识面,设计师要领导和协调整个项目的技术行为,负责建立各种总体框架结构视图。

· GIS 设计员(GIS Designer)。按照系统需求,负责定义一个

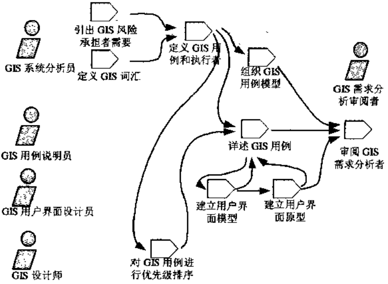


图 10—10 GIS 需求分析工作流程

或多个包(子系统),负责定义包或子系统中的类以及每个类的功能、操作、属性以及类间的关系。

· 数据库设计员(Database Designer)。负责 GIS 中数据库的设计。

· 结构和设计审阅员(Architecture reviewer and design reviewer)。对分析和设计的成果进行审阅。

——成果。GIS 分析与设计包括以下成果:

· GIS 设计模型(Design model)。系统建造的蓝图

· GIS 软件结构文档(Software architecture document),记录各种系统结构视图。

——工作流程。GIS 分析和设计的工作流程如图 10—11 所示。

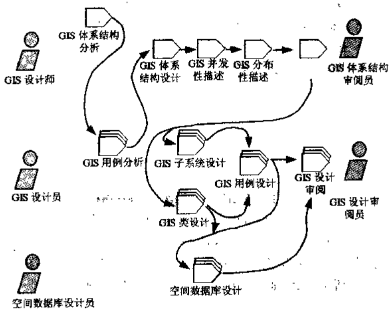


图 10—11 GIS 分析、设计工作流程

GIS 设计可以分为体系结构设计(Architectural Design)和详细设计(Detailed Design)两部分。体系结构设计是高层设计(High Level Design),其任务是定义包(或子系统),包括包之间的依赖性和主要通信机制。我们希望得到尽可能简单和清晰的结构,各部分之间的依赖尽可能少,并尽可能减少双向的依赖关系。详细设计的任务是细化包的内容,使编程人员得到所有类的一个足够清晰的描述。同时使用 UML 中的动态模型,描述特定情况下这些类的实例之间的行为。

——体系结构设计(Architectural Design)。一个设计良好的系统结构是系统可扩充和可变更的基础。包实际上是一些类的集合。类图有助于用户从技术逻辑中分离出应用逻辑(领域类),从而减少它们之间的依赖性。这就是软件结构设计强调的模块间的高聚合、低耦合的原则。在工具型 GIS 中,存在以下包:

- GIS 对象包。包括来自 GIS 用例模型的各种对象包。
- 数据库接口包。为 GIS 对象提供永久存储。
- 主控包。提供系统界面和消息响应的对象包。

这些包之间的关系如图 10—12 所示。

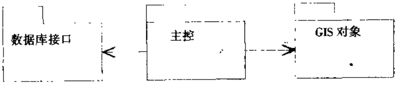


图 10—12 包图表示工具型 GIS 体系结构设计

——详细设计(Detailed Design)。GIS 详细设计的目的是通过创建新的类图、顺序图等,扩展和细化分析阶段"素描"的 GIS 对象类。从技术层次上对系统进行更详细的描述。用类图(如图 10—13 所示)来详细描述包所包含的类,用顺序图(如图 10—14

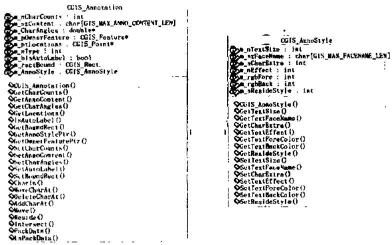


图 10—13 GIS 中的地图注记类图(局部)

所示)来展示 GIS 系统中每个用例在技术上如何实现。还可以用其他图来从不同侧面对分析阶段的模型进行细化。比如用状态图

可以描述单个对象在整个系统中的变化细节等。

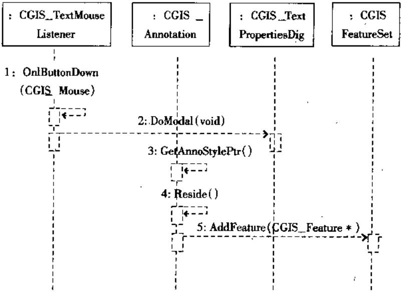


图 10—14 增加说明性注记顺序图

(4)GIS 系统实施(Implementation)。实施阶段是按照定义的子系统对类进行编程的过程。实施阶段还要完成系统的集成以及各组件的单元测试。在实现阶段中,可以选取类图、状态图或动态图(顺序图、合作图、活动图)来辅助编程。

编码期间也可能会发现设计模型的缺陷,这时需要开发者修改设计模型。修改设计模型时一定要保持设计模型与编码的一致性,以便将来易于维护。

早期的工具型 GIS 普遍采用基于 C/C++的编程语言进行编码,被称为第一代面向对象的工具型 GIS,这种 GIS 软件具有一系列的局限性,主要有系统的部件很难共享(C++的二进制组件很难共享,一般都共享源码),不能在不重新编译的情况下更新 C++组件,缺少好的建模语言和工具,以及合适的界面制作和客户

化工具，为了避开这些问题，近年来很多软件工程师们转向基于组件的开发方式。

软件工程组件技术的成功之处在于它用很具体的方法实现了软件工程中被普遍接受的面向对象思想。组件封装了经过精心定义的界面和方法，以及每个对象实例特有的数据。这样大大促进了结构化和安全的系统开发。组件技术方便了软件的再利用，因为它们是一个个自成一体的模块，可以很容易地组装成更大的系统。它们也支持继承性和多态性。组件可在别的组件里通过对对象的状态和行为的引用来对已有代码重新利用，即所谓继承。组件很容易地处理多态性，即对不同对象的相同操作定义不同的实现方法。多态性好处的一个例子是 GIS 可以有一个通用的对象创建组件，根据对象的类型进行不同的创建处理。

对 GIS 来说，对象/组件模型优于单纯的对象模型。因为它提供了一个可以让任何人扩展的数据框架。用户可使用 GIS 系统开发商同样的技术对数据和系统进行扩展。对用户而言，用户对象和 GIS 软件提供的对象没有本质的区别。

基于组件/对象方式的 GIS 设计和实施将很快变成新的规范。因为：

• 对象组件是符合标准的——界面的准确定义，或组件间通讯协议都有明确的标准。

- 可扩展性——因为组件是标准的，任何人(GIS 开发商、最终用户或第三方)都可以创建或扩展已有的 GIS 组件。新的组件可以在已有的组件上建立。
- 运行时扩展——现代操作系统，Internet 浏览器，以及大多数流行应用(e. g. Microsoft Office)都提供对系统运行时扩展的支持，由于组件只有在运行时才调入，可以最少地使用资源。
- 独立于编程语言——特别是对 COM 对象组件，它可以用任何现代编程语言创建。很多集成开发环境(IDE)提供

COM 和 Java 组件的建立和使用。

- 对象行为封装在数据模型中——在对象/组件模型中对象的行为有原来的用户应用中移到数据模型中，这对于提高数据的安全性和一致性非常重要。
- 强大的设计分析工具——有很多支持 COM 的集成开发环境(IDE)、CASE 工具，以及专业组件工具箱。GIS 软件开发者不再需要在其内核里单独开发诸如表格界面，报表生成器，统计图功能，而是可以采取“拿来主义”，将所有这些可以丰富 GIS 功能和表现力的第三方组件集成到系统中来。

——工作人员。在 GIS 的系统实施中，包括以下工作人员：

- GIS 实施人员(GIS Implementer)。开发组件并执行单元测试。
- GIS 系统集成员(GIS System Integrator)。对 GIS 各组件进行集成。
- GIS 设计师(GIS Architect)。定义 GIS 实施模型的结构。
- 代码审阅员(Code Reviewer)。检查代码的质量和一致性。

——成果。在 GIS 的系统实施中，包括以下成果：

- GIS 子系统(Subsystems)。组成 GIS 系统实施模型的各个部分，例如地图采集子系统，遥感更新子系统。
- 组件(Component)。软件代码(源代码、二进制代码或可执行代码)和一些信息文件，例如 Readme。组件也可以是其他组件的集成，例如由一些可执行代码组成的应用也是组件。图 10—15 显示地图注记实施模型的组件图。
- 集成计划(Integration build plan)。定义组件和子系统实施的顺序，定义集成规范。

——工作流程。IS 系统实施阶段的工作流程如图 10—16 所示。

(5)GIS 系统测试(Test)。完成编码后，需要对软件进行测试，它通常包括：

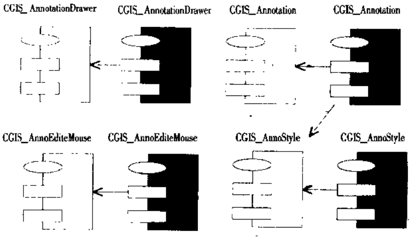


图 10－15　地图注记组件图(部分)

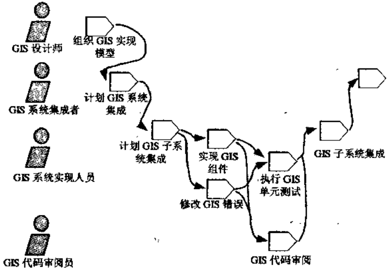


图 10—16　GIS 系统实现工作流程

- 单元测试(Unit test)

单元测试对单独的类或一组类进行测试，测试过程可以使用类图和类的规格说明作为指南。单元测试通常情况下是一种白盒测试，它由封装在类中的操作和类的状态行为所驱动。测试的内容包括类的接口(验证进出类对象的数据流)、类的数据结构(临时存储的数据在算法执行的整个过程是否维持完整性)、边界条件(类在极限或严格的情形下是否能正确执行)。单元测试必须对控制结构中所有的路径都要走遍，以保证类中所有的语句都至少执行一次，还要对处理错误的路径进行测试。

- 集成测试(Integration test)

完成单元测试的每一个类即使工作完好，也不意味着把它们放到一起就能正常工作。由于类之间通过接口互相影响，在一个类中的不确定性会很容易蔓延到整个子系统，如果使用了全局数据结构可能也会存在问题。集成测试对各组件或子系统的合作情况进行测试，从而发现接口有关的问题。测试过程使用组件图和合作图作为指南。

- 系统测试(System test)

在系统测试中，使用用例图，以检验所开发的系统是否满足用例图所描述的要求，主要是对系统的功能完整性和性能进行测试。系统测试通常由传统的黑盒测试方法驱动。

- 接收测试(Acceptance test)

接收测试则由最终用户进行，为最后的软件部署做准备。由于我们不可能预见用户实际上如何使用 GIS，也不可能让每一个用户都进行接收测试，因此一般厂商都使用被称作 alpha 和 beta 测试的过程来进行接收测试。Alpha 测试是由一个用户在开发者的场所进行，GIS 开发人员在用户的“指导”下进行测试，开发人员负责记录错误和使用中出现的问题。Beta 测试是由 GIS 的最终用户在一个或多个用户场所进行，开发人员一般不在现场。用户记录测试中遇到的

问题，并定期提交给开发者。开发者进行修改后，就准备向用户发布最终的产品。

——工作人员。在 GIS 的系统测试中，包括以下工作人员：

- 测试设计员(Test Designer)。负责测试的计划、设计、实施以及评价，建立测试计划和测试模型，实施测试过程评价测试范围、结果和有效性等。
- 系统测试员(System Tester)。执行 GIS 系统测试。
- 性能测试员(Performance Tester)。执行 GIS 性能测试。
- 集成测试员(Integration Tester)。执行 GIS 集成测试。

——成果。在 GIS 的系统测试中，包括以下成果：

- 测试计划(Test Plan)。测试的目的和目标，额外的软件，环境和资源，确定需要测试的方面和测试方法。
- 测试模型(Test Model)。测试用例、测试过程和测试脚本，测试描述、期望的结果等。
- 测试子系统和组件。
- 测试包和类。
- 错误(Defects)。测试失败的结果。

——工作流程。GIS 的系统测试工作流程如图 10—17 所示：

(6)GIS 系统部署(Deployment)。系统部署决定于软件的类型和特定的业务领域，工具型 GIS 系统部署一般包括：

——工作人员。在 GIS 的系统部署中，包括以下工作人员：

- 部署管理员(Deployment Manager)。负责对部署的计划、组织。
- 部署实施员(Implementer)。制作系统的安装软件。
- 技术文档编写员(Technical Writer)。编写用户手册，安装指南等。
- 课程规划员(Course Developer)。制作培训材料。

——成果。在 GIS 的系统部署包括以下成果：

- 部署计划

• 用户手册

• 培训计划——幻灯、在线教程等

——工作流程。

• 制作一个完全的外部版本

• 对软件打包

• 软件安装

• 培训用户

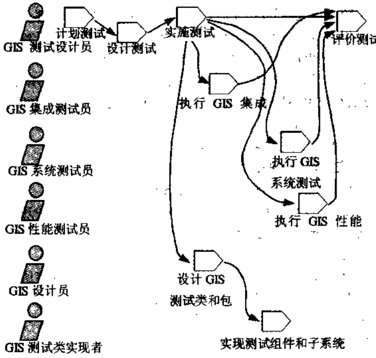


图 10—17 GIS 测试工作流程

§10.3 应用型GIS的开发和管理

应用型GIS是一种面向特定用户、解决特定领域问题、具有明确应用目标的地理信息系统，它通常是在工具型GIS基础上建立的，所选用的开发语言和GIS功能都依赖于工具型GIS，其技术比较成熟，风险较低。因此，应用型GIS开发的重点并不在于对基本功能的设计和实现，而是根据需求分析的结果，选择合适的地理信息系统平台，进行具体化的开发，以满足特定用户的需要。

应用型GIS的需求来源于不同行业的用户，可能一开始需求并不明确，但用户往往又要求很快地拿出成果，因此开发者往往使用快速应用开发的方法。由于本书的侧重点在于工具型GIS的理论、方法和技术，对于实际开发应用型GIS方法，还与采用的工具型GIS平台、二次开发语言有关，这一节只是简单地阐述应用型GIS的组织管理和基本开发过程。

1. 应用型GIS的组织管理

GIS的开发是一项十分复杂的系统工程，不仅涉及到GIS开发的技术人员，而且涉及到各个部门的用户，在开发过程中还需要经历可行性分析、系统分析、系统设计、测试、维护等多个步骤，如果没有有效的组织管理手段是难以成功的。

(1)机构的建立。为了能顺利地建立应用型GIS，并使所建立的GIS发挥出应有的效益，必须建立新的机构和部门。那种以为只要购买了计算机硬件和一些软件，组织一些人进行几个月的开发，然后就能立即得到收益的想法是不能如愿的。

GIS建立和GIS应用所需的人员是不同的，所建立的新机构主要是为了GIS的应用，比GIS建立所需的人员要少。但仍难免要调入一些GIS技术人员，和对原有人员进行新技术培训，以便能从技术上支持系统的正常运行。

新的GIS机构的日常工作主要是设备和软件的正常运行和

维护，数据的采集和更新，用户所需产品（如数字地图、图表等）的生成输出，等等。

（2）领导的支持。GIS 的开发与应用必须得到具有远见卓识的领导的支持，并直接参与 GIS 开发的组织工作。许多部门在引进 GIS 后都能以新技术、新方法迅速解决传统空间数据分析中很难解决的问题，人们以这些积极因素作为发展和应用 GIS 的重要理由，但应该认识到新技术也会带来新问题，当新技术不能很快实现人们所希望的目标时，领导层不应失去支持新技术的信心。

一种普通存在的现象是，在短期内，GIS 的处理结果往往比传统方法的费用高许多，其主要原因是数据采集工作是一项既费时又费钱的工作，在短期内不能充分发挥这些数据的效用。此时，领导者为了保持 GIS 的生命力，应继续支持和投资，使 GIS 不断完善，逐步达到回收投资的目的。

（3）经费的支持。经费的支持是关系到 GIS 成败的主要决定因素。除了在建立 GIS 的最初阶段的投资金额外，在应用 GIS 的各个阶段还要进行连续不断的投资。最初的投资主要用于硬软件购置、人员培训、机房建设等，持续的投资主要用于数据采集、设备维护、软件开发等。

硬件设备的投资在 GIS 总投资中占较大比重。因此，除了按预算金额提出设备清单外，还要考虑硬件使用的优先顺序。把 GIS 开发马上就绝对需要的设备和一段时间以后才绝对需要的设备作为优先和次优先项目购置。对于那些虽然需要，但使用不多的设备（如一些高质量的输出设备），可通过借用而省去大笔投资。

在购买 GIS 开发软件时，应选择功能强，价格合理的工具软件，并且要能满足所需开发 GIS 的目的和要求，当然还要与所选择的硬件设备相配套。

（4）人员的组织。今天，计算机技术和信息技术的进步已经大大领先于人脑的进化和对人脑进行教育的速度。成功的软件项目中最重要的因素并不是使用的工具，而是人。在我国，对于富有创

造性的软件开发人才的实际重视有时与管理者的表态不符，这可能和长期以来非创造性的劳动力成本低密切相关。一些组织在发现由于人才外流使他们越来越难以应付需求日益增长、并且日益复杂的软件开发时，才意识到如何吸引、培养、鼓励和留住所需的人才，并给他们提供良好的工作环境是多么的重要。

在 GIS 的开发与应用中，需要不同层次和类型的人才，并且 GIS 的开发和 GIS 的应用所需的人员的数量和类型是不同的。许多需要建立 GIS 的部门并不具备开发 GIS 的能力，因为 GIS 的开发需要有 GIS 专家和一批 GIS 开发人员，这时，要么调人有关人才自行开发，要么委托他人开发，自己进行验收和应用。在 GIS 的开发中，通常需要以下人员：

· 组织管理人员

GIS 的开发是一项大型工程，组织管理者是极为重要的。由于 GIS 的涉及面很广，如资金的使用、硬件的配置、人员的调配、数据的来源、应用的权限等，所以必须有领导者的参加，同时还需要一些具有 GIS 知识的公关协调人员。

· GIS 专家

GIS 专家应具有测绘学、地理学、计算机科学等方面的知识，精通 GIS 技术，并且具有较强的业务组织能力。他们要负责系统设计、系统分析等工作，能解决 GIS 开发中的难点，并对系统的开发过程进行控制。

· GIS 的软件开发人员

他们应具有 GIS 方面的知识，以及图形、图像处理方面的能力，尤其是能熟练地利用 GIS 的开发工具进行系统功能的开发。他们负责 GIS 的具体开发工作。

· 数据采集人员

数据是 GIS 的血液。数据采集人员应具有一定的计算机基础和 GIS 的常识，能熟练地进行 GIS 的数据采集工作。

· 维护人员

作为一个大型的软件开发工程，还需要有软硬件维护方面的人员，以及软件文档管理人员。

除此以外，GIS 的开发必须有用户的积极参与，这样才能开发出满足用户要求的 GIS。

2. 应用型 GIS 的开发过程

应用型 GIS 的开发过程通常分为以下几个阶段：可行性研究、系统总体设计、系统详细设计、系统实现、系统运行和维护，并且应在一个阶段的工作完成后再进行下一阶段的工作，每一阶段的工作有其明确的完成标志(里程碑)，这样才能对应用型 GIS 的开发过程进行严格而有效的管理。应用型 GIS 各开发阶段的主要工作可简述如下：

(1)可行性研究。可行性研究是应用型 GIS 开发的第一个阶段。其目的是调查应用型 GIS 的项目开发能够解决哪些主要问题，确定应用型 GIS 所要完成的任务，需要什么样的硬件、软件、数据库、通信设备、数据、人员等，对成本和效益进行预测，有什么样的风险，以及建成后所产生的效益，从而分析建立应用型 GIS 的必要性和可能性。调查的结果形成可行性研究文档提交高层管理部门，以作出"进行/不进行"的决策。

可行性研究主要包括：

· 经济可行性

评估收入或收益是否能够超过成本(应用于国防的 GIS 项目除外)。

GIS 是一项耗资极大的系统工程，除了在建立 GIS 最初的投资外，在应用的各个阶段都需要连续不断的投入。资金的投入往往决定了系统的建设规模。GIS 开发的费用主要用于：硬件、软件、技术开发、人员培训、数据采集、系统维护、消耗材料等。对资金在硬件、软件、数据和管理等几个方面的投入应有合理的分配。通过经费预算来确定是否有资金保证 GIS 的建设。

GIS 建设的投资较大，短期内效益往往并不明显，其效益通常

从投资回收期、效益/费用比、节省人力物力、减轻劳动强度、提高工作效率、提高管理水平、数据处理的及时准确、提供决策依据、应用服务项目等各个方面进行分析。

・技术可行性

研究应用型 GIS 的系统功能、性能和约束(是否有足够的技术力量和软硬件设备、相关技术的成熟程度以及是否掌握了 GIS 的关键技术)。

建设 GIS 需要各种层次的专业技术人员，而 GIS 的运行对用户也有技能方面的需求，因此需对技术力量的数量、结构进行分析，确定是否满足建设 GIS 的要求。GIS 的技术力量不仅要考虑数量，更重要的是要考虑质量，以及对用户培训和可以达到的水平。在 GIS 建成以后，用户的技术能力和管理水平直接影响到 GIS 效益的发挥，例如，作为 GIS 用户应注意培养 GIS 硬软件、数据输入、操作应用、产品输出、组织管理等方面的人才。

・法律可行性

确定可能产生的任何侵权、妨碍或责任。

・社会可行性

领导是否支持、组织管理和规章制度是否健全。

・进度可行性

评价应用型 GIS 项目能否在规定的时间期限内完成。

(2)需求分析。在用户需求分析中，需要弄清用户现行业务的运作过程、GIS 需要解决的问题等，据此确定系统的轮廓和目标。

对用户现行业务的运作过程进行详细分析是对 GIS 的建立非常重要的，无论用户现行业务的运作是人工进行的，或是由计算机实现的，都要了解业务的运作流程、信息的处理方式、资料的使用形式、人员和设备的配置、经费的使用等情况。同时找出现行业务运作方式存在的问题，逐步明确 GIS 应解决的问题，初步确定系统的目标。

用户需求调查的方法可采用访问、座谈、填表、抽样、查阅资

料、深入现场、与用户一起工作等方法。

(3)系统总体设计。

·确定系统的具体目标

在对用户现有工作流程深入调查的基础上，应明确所建应用型GIS的目标、规模、界限、主要功能、组织机构、业务流程、数据流程、数据存贮、存在问题等，以确定所建系统的具体目标。

根据所建系统对数据的使用、加工和处理过程，获得系统的数据流图。GIS是处理空间数据的，它的运行需涉及大量的图形、表格、文档等数据，而数据流图代表了数据操作的逻辑模型，是对业务作业过程的一种有效反映，在GIS建设中具有重要作用。

对数据流图中出现的所有空间数据、属性数据进行描述和定义，就形成了数据字典。数据字典应对数据流、文件、数据项、加工者名称、组成、取值范围、数据类型等进行详细的说明。

·确定系统的硬软件配置

硬件在GIS的投资中占有不小的比重，应选择质量好，并有软件支持的设备。硬件设备主要包括计算机、存贮设备、数字化仪、绘图机、扫描仪、打印机及其他外部设备。最后应绘出硬件设备配置图。

对软件应选择功能强、价格合理、有升级能力，并符合硬件要求的产品。

·子系统及其功能设计

一个大型的GIS通常包含若干个子系统，依据功能的聚散程度、运行过程的邻近性、行政管理机构的设置等来划分子系统，并确定子系统的功能和子系统之间的接口，然后给出系统的总体结构图，并将子系统的功能、接口方式、权限设置等记录在有关文档中。

·数据库设计

GIS数据库设计的重点是数据库的结构。需要确定空间数据库和属性数据库的结构，以及空间数据库和属性数据的连接关系。

当选择了 GIS 的开发平台软件(如 ARC /INFO)时,应根据其提供的数据库管理系统进行设计。

(4)系统详细设计:

·功能模块设计

在确定了子系统的划分后需进一步确定子系统由哪些模块构成,以及模块之间的接口、模块实现的方案、及其他技术细节。在模块设计完毕后,应绘出各模块的结构图,以详细描述各模块的功能、内容、接口等。

在模块划分和设计时,应使各模块功能特点清楚、各模块应尽量相互独立、模块之间关系清晰、易于实现和操作等。

·数字化方案设计

根据 GIS 对不同数据的要求,确定数字化方案。例如数字化时的要素分层、关系处理、精度要求、作业步骤等。从地图上采集数据的数字化工作通常包括:原图检查与准备、数据采集、数据文件生成、格式转换、数据编辑、属性表建立、属性匹配、注记输入、对照检查、屏幕修改、数据入库等。

在数字化方案设计时还应考虑数字化硬软件的影响,并应制订空间数据的标准和规范(尽量采用国家标准和规范),同时还需制定数据质量控制与检查的方法。

·输入输出设计

确定哪些设备进行输入,哪些设备进行输出,输出哪些内容,输入输出数据的格式(如能接收哪些格式的数据,输出的数据产品是什么格式,中间转换文件的格式等),以及接口软件的设计等。

·用户界面设计

用户界面在 GIS 中的地位越来越高,不仅要方便用户的操作,而且要尽可能与当前流行的操作界面一致,使用户能用尽可能少的时间和精力用于操作的学习。在界面设计时,应对图面布局形式、图面布局内容、色调搭配、菜单形式、对话方式等加以说明。用户界面一定要简明,并以用户熟悉的专业或自然语言来表达 GIS

的功能和操作，使用户能很快地学会操作，操作方法简便直观，用户能迅速而自然地找到要进行的操作，等等。

(5)系统的实现。

·硬软件设备的安装与调试

这是系统实现的基础，GIS 必定是在一定的硬软件支持下建立的。

·数据准备和数据库建立

GIS 所涉及的数据种类较多，数据结构复杂。数据的准备与采集是一项工作量巨大、工作内容复杂的任务，必须严格按照数字化作业方案的要求进行数据输入和建库的工作。数据输入人员应经过相应的培训。

·软件的编制与调试

软件的编制与调试是对 GIS 详细设计结果的具体实现，通常是在所选用的工具型 GIS 所提供的开发环境下进行的。要在源程序编制中遵守软件研制的基本原则，如格式、命名、注释等，并且应对各源程序单元进行测试，验证模块功能和接口与详细设计说明的一致性。

·系统联调与试验

当各子系统的软件编制与调试完成后，需把各部分组织起来进行联调，主要是发现与接口有关的问题。同时应拿实际的数据对系统进行试验，检查系统是否达到设计的要求。

·测试与验收

在软件开发过程中，测试是必不可少的、经常性的工作之一。此处所讲的测试主要是指 GIS 系统验收过程中的测试。通常由验收委员会组成专门的测试小组，小组中应有领导、专家、研制单位或部门可有少量人员参加，以便于测试与咨询，必要时可请第三方参加。测试主要包括：测试前检查、系统演示、功能测试、性能测试、强度测试等，最后应写出测试报告。

GIS 系统的验收必须履行正式手续，需组织验收委员会进行

验收。验收的主要工作是:审查验收申请、制定验收计划、验收测试、验收评审、形成验收报告等。最后移交系统。

· 操作人员培训

应提前对用户的操作人员进行培训。对于用户中的一般人员及领导人员,也需进行一定的宣传,使其对新建的GIS有所了解。以便更好地发挥GIS的效益。

(6)系统的运行与维护。

· GIS的使用

在业务工作中使用GIS以产生效益。如果所建的GIS只能用于演示,那就是失败的。

· 数据更新

只有持续地进行数据更新,才能保持GIS数据的现势性,GIS才能正常地发挥作用。

· 软件维护

软件维护是在软件产品交付使用后,为纠正错误、改进性能等而进行的修改。软件维护分为:适应性维护、纠错性维护、完善性维护和预防性维护,应建立维护小组及主管人员。每次维护工作都应审批和验收,并撰写相应文档。

· 硬件维护

主要包括机器设备的日常管理和维护,以及对硬件设备的更新。

第十一章　RS、GPS 与 GIS 的结合

§11.1　遥感(RS)与 GIS 的结合

1. 遥感(RS)的基本原理与技术

(1)遥感的基本概念。遥感是一种远离目标，通过非直接接触而判定、测量并分析目标性质的技术。

对目标进行信息采集主要是利用了目标反射或辐射的电磁波，此外重力和磁也作为信息采集对象而加以利用，这些都包含在广义的遥感之中。

接收从目标中反射或辐射的电磁波的装置叫作遥感器(remote sensor)，照相机及扫描仪等即属于此类。此外，搭载这些遥感器的移动体叫做遥感平台(platform)，如现在使用的飞机及人造卫星等。遥感这一词汇是 60 年代在美国创造的技术用语，它是用来综合以前所使用的摄影测量、像片判读、地质摄影而提出的。特别是 1972 年，随着第一颗地球观测卫星 Landsat 的发射成功而迅速得到普及。

利用遥感技术，通过观测电磁波，从而判读和分析地表的目标及现象，是利用了物体的电磁波特性，即"一切物体，由于其种类及环境条件不同，因而具有反射或辐射不同波长的电磁波的特性"。所以遥感也可以说是一种利用物体反射或辐射电磁波的固有特性，通过观测电磁波达到识别物体存在的环境条件的技术。

根据可利用的电磁波的光谱段，遥感可分为可见光与反射红外遥感、热红外遥感、微波遥感三种类型。

在可见光与反射红外遥感中，所观测的电磁波的辐射源是太

阳。这种遥感的数据对地表目标物的反射率有很大的依赖性，即根据反射率的差异可获得有关目标物的信息。这里，激光雷达是个例外，它的辐射源是其装置本身。

在热红外遥感中，所观测的电磁波的辐射源是目标物。

在微波遥感中，所观测的电磁波的辐射源有目标物(被动)和雷达(主动)两种。在被动微波遥感中，是观测目标物的微波辐射，在主动微波遥感中，是观测目标对雷达发射的微波信号的散射强度。

(2)航空遥感技术。航空遥感是以飞机或气球作为工作平台进行成像或扫描的一种遥感方式。在工作平台上装有各种传感器，并对测区地物的电磁波信息进行收集、处理，最后获得图像或数据，并用于各种目的，这样的过程即称为航空遥感。

由于航空遥感具有成像比例尺大、分辨率高、几何纠正准确等优点，因此是重要的遥感手段，在GIS中也有重要的应用。

航空遥感的平台不仅包括飞机、气球，而且还包括有人驾驶和无人驾驶的遥控飞机。其传感器的类型种类繁多，不仅有光学摄影的航摄仪，还有多波段扫描仪、红外扫描仪以及侧视雷达等。成像的种类也多种多样，从可见光的黑白全色摄影、彩色摄影，到紫外、红外、微波等波段，从而可以收集不同宽度波段和特定波长的电磁波信息。

现在的航空遥感，除以感光胶片作为传统记录外，还可进行磁带记录，把获取的紫外、红外、微波等信息，转换成图像和数字信息，以便于计算机对遥感信息的进一步处理。

(3)航天遥感技术。从航空遥感发展到航天遥感，是遥感技术的重大突破。人造卫星给多种传感器提供了离地面更高的工作平台，使传感器具有更广阔的视野，居高临下，俯视一切。卫星遥感还可不受国界和地形的限制，可对全球作连续的观测。

遥感卫星根据其轨道及搭载的遥感器不同而有不同的特征。从赤道上空约3600Km的静止轨道长期观测是以气象观测为主要

目的的；从高度约 700Km 的极轨道观测地球表面的 Landsat 卫星是以陆地观测为主要目的的；从高度约 850Km 的极轨道观测云及水陆的各种现象的气象卫星 NOAA 是以气象及海洋观测为主要目的的，但也应用于了解全球性陆地植被的分布等。

遥感卫星的主要功能有，记录准确的遥感器的位置，可靠地获取数据以及将获取的数据传送到地面站。所以，利用卫星上的姿态控制系统测量位置，进行观测仪器的监视（包括地面跟踪、记录的控制）和数据传送是非常重要的。

2. 遥感数据的计算机处理

（1）遥感数据。象普通像片那样，灰度和颜色连续变化的图像叫模拟图像；而把模拟图像分割成同样形状的小单元，以各小单元的平均亮度值或中心部分的亮度值作为该单元的亮度值进行数字化的图像叫做数字图像。把前一部分的空间离散化处理叫做采样，而后一部分的亮度值的离散化处理叫量化，两个过程合起来叫做图像的数字化。

遥感数据是用遥感器探测来自地表的电磁波，通过采样及量化后获得的数字化数据。

遥感中所采用的电磁波的波长范围是从紫外线到微波。用遥感器接收的电磁波是按每个特定的波长段进行数字化的。通过这一过程可以得到把具有连续波长分布的电磁波分解为若干个波长段（称通道或波段）的数据。这种由多个波长段构成的数据叫多通道数据或多波段数据或多光谱数据。

在多波段数据中，各波段的电磁波是用不同的探测元件探测的。因为从地表到达受光面上排列的各探测元件的路径有微小的不同，所以各探测元件捕捉到的地表位置存在着细小的差异。校正这种差异，把各通道的数据重合起来的过程叫通道间的配准 。

多波段图像具有位置和光谱信息，其数据格式根据二维空间的像元配置中如何存贮各种波段的信息而分成三类：

①BSQ 方式——各波段的二维图像数据按波段顺序排列。即

{[(像元号顺序),行号顺序],波段顺序}

②BIL 方式——每一行代表一个波段的光谱值,然后按波段顺序排列行,最后对各行进行重复。即{[(像元号顺序),波段顺序],行号顺序}

③BIP 方式——在一行中,每个像元按光谱波段次序进行排列,对该行的全部像元进行这种波段次序排列,最后对各行进行重复。即[(波段顺序,像元号顺序),行号顺序]。

在遥感数据中,除图像信息以外,还附带有各种注记信息。这是提供数据的机构在进行数据分发时,用注记的形式来提供说明。

遥感卫星的观测数据通常可以通过运行该卫星的机构或世界各地的接收站、数据分发中心等有偿或无偿地获得。

(2)图像校正。由于大气层的存在,也由于传感器内检测器性能的差异等情况,使进入传感器的辐射值发生畸变,引起图形模糊,对比度下降等。另一方面,由于卫星飞行时的姿态变化及地球形状、地表形态等因素的影响,图像中地物目标的几何位置也会发生畸变。

图像校正就是指从具有畸变的图像中消除畸变的处理过程。图像校正分为辐射量校正和几何校正。

利用遥感器观测目标物的辐射或反射的电磁能量时,从遥感器得到的测量值与目标物的光谱反射率或光谱辐射亮度等物理量是不一致的。这是因为测量值中包含太阳位置及角度条件、薄雾及霭等大气条件所引起的失真。为了正确评价目标物的反射特性及辐射特性,必须消除这些失真。消除图像数据中依附在辐射亮度中的各种失真的过程叫辐射量校正。辐射量校正包括由遥感器的灵敏度特性为起因的畸变校正,由太阳高度及地形等为起因的畸变校正、大气校正等。

图像中所包含的几何畸变可以表示为图像上各像元的位置坐标与地图坐标系中的目标地物坐标的差异。几何畸变的原因可大致分为4类:

①遥感器的内部畸变:由遥感器结构引起的畸变。

②遥感器的外部畸变:由图像投影方式的几何学引起的畸变。它可以进一步分为平台引起的畸变和目标物(地球的自转等)引起的畸变。

③图像投影面的选取法引起的畸变:由于图像投影面的选取法(图像坐标系的定义方式)不同,几何畸变的表现也不同。

④由地图投影引起的畸变:根据采用的地图投影的不同,几何畸变的表现也不同。

图像的几何校正是指从具有几何畸变的图像中消除畸变的过程。也可以说是定量地确定图像上的像元坐标(图像坐标)与目标物的地理坐标的对应关系(坐标变换式)。

图像的几何校正步骤大致如下:

①确定校正方法:考虑到图像中所含的几何畸变的性质及可应用于校正的数据确定校正的方法。如系统性校正、非系统性校正、复合校正等。

②确定校正式:确定校正式(图像坐标和地图坐标的变换式等)的结构,根据控制点数据等求出校正式的参数。

③验证校正方法、校正式的有效性:检查几何畸变能否充分得到校正,探讨校正式的有效性。当判断为无效时,则对新的校正式(校正方法)进行探讨,或对校正中所用的数据进行修改。

④重采样、内插:为了使校正后的输出图像的配置与输入图像相对应,利用②中所采用的校正式,对输入图像的图像数据重新排列。在重采样中,由于所计算的对应位置的坐标不是整数值,所以必须通过对周围像元值进行内插来求出新的像元值。

在几何校正的最后阶段,必须根据几何畸变的校正式对输入图像进行重采样,以便输出无畸变的图像。重采样的方法有以下两种:

①对输入图像的各个像元在变换后的输出图像坐标系上的相应位置进行计算,把各个像元的数据投影到该位置上。

②对输出图像的各个像元在图像坐标系的相应位置进行逆运算，求出该位置上的像元数据。该方法是通常采用的方法。

在通常的几何校正过程中，是在校正后的输出地图坐标系上设定正方格，求出格点上对应的图像数据。可是，在该输出图像坐标系的格点上所对应的输入图像坐标通常不是整数值，所以必须用输入图像上周围点的像元值对所求点的像元值进行内插来求出。用于几何校正的主要内插方法有以下 3 种：

①最邻近内插法。以距内插点最近的观测点的像元值为所求的像元值，该方法最大可产生 1/2 像元的位置误差，但优点是，不破坏原来的像元值，处理速度快。

②双线性内插法。使用内插点周围的 4 个观测点的像元值，对所求的像元值进行线性内插。该方法的缺点是破坏了原来的数据，但具有平均化的滤波效果。

③3 次卷积内插法。使用内插点周围的 16 个观测点的像元值，用 3 次卷积函数对所求像元值进行内插。该方法的缺点是破坏了原来的数据，但具有图像的均衡化和清晰化的效果，可得到较高的图像质量。

(3)图像变换。当分析图像数据时，为了使分析者能容易确切地识别图像内容，必须按照分析目的对图像数据进行加工，这一处理过程叫图像增强。典型的图像增强有灰度变换、彩色合成、色彩空间变换等。

另一方面，为了进行图像判读及分析处理，需要从图像数据中求出有益于分析的判读标志及统计量等各种参数。把图像所具有的性质进行定量化的处理过程叫特征提取。实际上，图像增强和特征提取多用于同一处理中，前者是以人工判读为前提，重点放在显示技术上，后者的特征则是处理的定量性。

①灰度变换。灰度变换是指通过对图像的灰度信息进行变换处理，而达到增强图像的目的。它是通过把原图的灰度值 x 用函数 f 变换为值 y 而进行的。即：

$$y=f(x) \tag{11—1}$$

典型的灰度变换方法有：线性变换、分段线性变换、三角波变换、连续函数变换、局部灰度变换等。

②直方图变换。直方图变换是使输入图像的灰度值的频率分布（直方图）与所希望的直方图形状一致而变换灰度值的方法。

典型的直方图变换分为直方图均衡化和直方图正态化。直方图均衡是指把原图像的直方图变换为各灰度值频率固定的直方图；直方图正态化是指将图像的直方图变换成正态分布的形状。

③图像的彩色显示。遥感数据是直接从遥感器得到的数字数据的罗列。为了使其内容直观易懂，常需进行彩色显示。彩色显示有两种方法，一是把多个波段的图像分别赋予一种原色而进行显色的合成法；另一种是对黑白图像的灰阶赋予颜色的假彩色显示法。

④图像间运算。在以同一区域为目标的不同图像间进行运算的处理过程称为图像间运算。包括多光谱图像的波段间运算和不同时期观测的图像间运算等。运算的结果生成新的图像数据。

图像间的运算分为算术运算（加减乘除及其组合）和逻辑运算（逻辑和、逻辑积等）。

⑤主成份分析。主成份分析着眼于变量之间的相互关系，在尽量不丢失信息的前提下，用几个综合性指标对多个测量值进行描述。在多光谱图像中，由于各波段的数据间存在相关的情况很多，通过主成份分析可以把图像中所含的大部分信息用假想的少数波段表示出来。这意味着信息几乎不丢失，但数据量可以减少。

⑥空间滤波。空间滤波是指对图像应用若干滤波函数而获得改进的输出图像的技术。其效果有噪声的消除，边缘及线的增强，图像的清晰化等。

滤波函数有低通滤波、高通滤波、带通滤波等。低通滤波用于仅让低频的空间频率成分通过而消除高频成份的场合，由于图像的噪声成份多数包含在高频成份中，所以可以用于噪声的消除。高

通滤波仅让高频成份通过，可应用于目标轮廓等的增强。高通滤波由于仅保留一定的频率成份，所以可以用于提取，消除每隔一定间隔出现的干涉条纹的噪声。

⑦纹理分析。纹理也称结构，是指存在于图像某一范围内的有规律排列的图案。在图像判读中使用纹理表示图像的均匀、细致、粗糙等现象。根据构成图案的要素的形状、密度、方向等纹理进行图像特征提取的处理叫做纹理分析。

(4)图像分类。将图像中所含的多个目标物区分开就是图像的分类。

首先根据应用的目的和图像的特性确定分类类别；然后找出这些类别的特征量；为了得到分类的基准，需要提取出各个类别的训练数据；为了测定总体特征，需对图像中不同的部分进行采样或对特征相似的像元进行归类；然后使用指定的分类基准对各像元进行分类；最后要把已知的训练数据及分类类别与分类结果进行比较，以确定分类的精度和可靠性。

分类的常用方法是：多级切割分类法、决策树分类法、最小距离分类法、最大似然比分类法、模糊分类法、专家系统分类法、神经网络分类法等。

3. RS 与 GIS 的结合

RS 与 GIS 的结合具有重要意义。GIS 的生命力将最终取决于其空间数据库的现势性，且这一问题随着 GIS 技术的成熟而显得更加突出。遥感数据是 GIS 的重要信息源和数据更新的手段。可以从遥感图像中快速而可靠地提取地面目标的空间和属性信息。我国卫星遥感地面站的实际工作经验证明，大约每 1－2 年 TM 图像可覆盖我国国土一次，其数据的现势性比常规的地图资料要好。因此利用遥感图像实施 GIS 空间数据的更新具有重要意义和价值。

同时，RS 与 GIS 的结合可以有效地改善遥感分析。从遥感技术的发展看，随着高分辨率和多种传感器遥感图像的问世，正面临

越来越高的应用要求，而GIS为达到这个要求提供了良好的技术环境。目前在如何利用GIS中的信息提高遥感图像的分类精度、引导对图像中的目标进行空间搜索和提取等方面都已有了深入的研究。

另外，RS与GIS的结合可以进一步加强GIS的空间分析功能。一方面现代遥感技术的多平台、多波段、多时相、周期性的实时采集可以获取直观的地表景观信息，不仅真实地反映了区域的地理环境信息，面且能动态地反映社会经济环境情况。另一方面，随着遥感技术的发展，遥感信息自动识别水平大大提高，为我们利用遥感信息进行空间分析提供了先进手段，也保证了GIS空间分析的现势性。因此，可以利用信息复合处理技术，与GIS中的空间信息和属性信息结合起来进行空间分析，可以获得更可靠和更广泛的分析结果。

(1)RS与GIS的结合方式。RS与GIS的结合方式通常有三种：

①分开但是平等的结合。各系统有不同的用户界面，不同的工具库和不同的数据库。RS与GIS之间只传送数据，如图11—1。

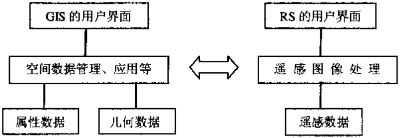


图11—1 RS与GIS的结合方式一

②表面无缝的结合。有统一的用户界面，不同的工具库和数据库。如图11—2。

③整体的结合。同一个用户界面、工具库和数据库，这是未来应采用的理想模式，如图11—3。

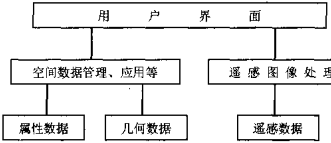


图 11—2 RS 与 GIS 的结合方式二

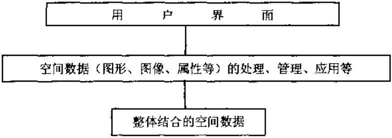


图 11—3 RS 与 GIS 的结合方式三

(2)RS 用于 GIS 地理数据库的快速更新。利用遥感图像处理技术实现数据快速更新的方法步骤如图 11—4 所示。

图像预处理的主要目的是根据资料及所需要的专题信息设法提高遥感图像的可分性。其技术方法有反差增强、边界增强、均衡化、比值组合处理(产生比值影像、差值影像)、色度空间转换等。通过图像预处理,便于提取其中一些可分性高的波段和目标。

自动识别分类处理是对预处理后的图像进行模糊分类、领域分类和纹理分类等,然后选择分类结果最好的一种。

实现数据快速自动更新的关键在于提高分类精度。如果分类结果不令人满意,还要重新选择波段和分类方法。另外,可采用人机交互判读、利用专题信息引入编译分类等方法来完成。目前,虽然以人工智能为特点的遥感图像判读专家系统有广阔的发展前景,但尚属试验阶段。而人机交互判读是当前在保证实用的前提

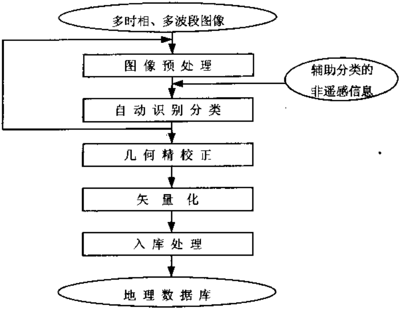


图 11—4 RS 用于 GIS 更新数据的方法步骤

下，提高遥感数据分类精度的一条切实可行和十分有效的途径。它可以很好地发挥人脑和电脑的优势。完成图像的局部放大、手动勾画或修改界线、标注或属性等任务。

几何精校正处理是为了与地图数据配准，取得投影方式一致，比例尺一致，便于入库存贮而进行的，也就是用与传统的图像同名控制点相对应的方法来实现。

通过遥感影像的正射纠正、投影转换、坐标变换等，遥感影像应与 GIS 的空间数据在同一数据基底下配准。配准精度应满足基底数据精度，对 1:25 万数据而言，配准误差不应大于图上 0.2 毫米。

为了更新 GIS 的矢量数据库，需要将遥感图像与 GIS 空间数据对应的图形以透明方式迭加，并发现和确定需要更新的内容。然后要将栅格数据进行矢量化处理，同时进行一些入库前的预处理。

如处理数据更新区域新数据与旧数据的关系，建立拓扑关系等。然后数据就可以按 GIS 指定的数据结构入库了。

航空摄影像片影像分辨率高，几何精度好，仍是更新 1:5 千、1:1 万、1:2.5 万、1:5 万、1:10 万比例尺地图的主要资料。随着航天遥感技术的发展，航天遥感图像的分辨率、几何精度、光谱分辨率的提高，航天遥感图像不仅可以用于 1:5 万比例尺地图的更新，还可用于 1:1 万、1:2.5 万比例尺地形图的更新。

用卫星影像获取各种地面要素的矢量信息，可选用美国陆地卫星的 TM 多光谱图像或法国的 SPOT 卫星图像。SPOT 卫星图像的产品有全色、多光谱和立体图像三种。全色图像地面分辨率为 10 米，可用于比例尺小于 1:5 万地图数据的更新。陆地卫星 TM 多光谱图像地面分辨率为 30 米，适用于比例尺小于 1:25 万地图数据的更新。针对不同物体各自的光谱特性，即物体不同，其吸收、反射和辐射电磁波的性能各不相同，当同时使用几个不同的光谱段获取影像时，不同物体就在不同光谱段的图像上被突出的表现出来。根据地物要素的光谱特性和利用卫星的光谱段，可以直接在卫星图像上判读各种地物要素。例如，植被在红外波段反射率最大，而水体在近红外波段辐射强烈吸收，因而在多光谱彩色图像上会呈现截然不同的颜色，有利于图像判读。

水系要素的更新：由于水体对近红外辐射强烈吸收，因而水体在 TM4 波段影像上明显地显示出来。利用纠正后的 TM4 波段影像，通过目视判读，确定兴趣区域，再利用常规自动化图像分割方法或选定种子点进行区域分割的半自动方法提取网状和面状水体的轮廓。轮廓确定之后，可在迭加的基础上，实施自动更新或采用对照提取结果手工编辑原始数据的方法进行更新。

道路要素的更新：新修的道路在 TM2 波段影像上反射强烈，目视判读容易，具有一定的形状特征，且辐射特性与邻域目标反差较大。可利用边缘提取算子、线状地物中心骨架确定方法、自动边缘检测方法确定道路的边缘或中心骨架，同样在与矢量信息迭加

的基础上实施自动或手工半自动要素更新。

居民地要素的更新：面状居民地在 TM2 波段影像上比较明显，因此面状居民地可利用 TM2 波段提取或采用数字化方法进行更新，轮廓提取的方法近似于水系要素轮廓的确定方法。居民地要素需要与道路要素进行迭加，并对道路与居民地的相互关系加以判断。

植被要素的更新：1:25 万比例尺的植被层变化较大，可利用植被在红外波段反射率最大的特性，直接在 TM2,3,4 三波段彩色合成影像上目视判读植被的边界并在屏幕上半自动确定分割边界，或采用多波段合成影像进行分类，最后获得植被要素的矢量数据。

其他要素的数据更新：同样需要根据地物的反射或辐射特性选择相应的遥感影像，并与要素的矢量数据配准迭加，更新方法同样可采用自动与半自动两种方式。

由遥感影像支持的 GIS 数据库的更新可以有两种途径：

①对 GIS 总体数据库的直接更新。这里存在访问权限问题，只能由高级用户或系统管理人员负责实施。对一种地物要素的数据进行的变化处理，需要同时更改与之联系的所有地物数据及其相关关系。

②处理结果的独立存贮。从遥感影像上提取的有关地物的相关信息，采用与总体数据库相同的数据结构加以存贮，作为一种新的数据源辅助 GIS 数据库的建立、维护与更新。

(3)利用 RS 进行 GIS 的空间分析。目前，一方面现代遥感技术的多平台、多波段、多时相、周期性的实时采集可以获取直观的地表景观信息，不仅真实地反映了区域的地理环境信息，而且能动态地反映社会经济环境情况。另一方面，随着遥感技术的发展，遥感信息自动识别水平大大提高，为我们利用遥感信息进行空间分析提供了先进手段，也保证了 GIS 空间分析的现势性。

下面以军事地理信息系统中利用遥感信息进行空间分析的几

个方法为例，对在GIS中如何应用RS进行空间分析作一简要介绍。

①利用遥感信息进行军事交通的分析。军事交通情况是部队机动的重要条件，它包括越野通行情况分析和正常道路(铁路、公路、航运等)的战时通行能力分析。其中越野通行是交通分析的主要内容。它指的是履带式车辆离路后影响其通行的因素和影响程度，这与自然地理环境中地形、水系、土质、植被有着密切的联系。利用遥感信息进行军事交通分析，实质上就是遥感信息与非遥感信息的综合分析。利用信息复合处理技术，将遥感图像信息中的地形信息与GIS中的DEM数据、坡度数据复合来辅助遥感图像的监督与非监督分类，确定最佳阈值，提高通行情况的分区精度。最终根据复合情况画定分区边界线，确定出易通行区、通行困难地区和不易通行地区。还可将上述军事交通专题信息，利用GIS中电子地图制图技术叠加到相应的地形图上，制作出现势性强的军事越野通行专题图供部队使用。另外，随着经济的发展，正常道路的变化也很大，对战时部队的机动影响也较大。利用航空像片或卫星影像可以对正常的军事通道，如铁路、公路进行更新和进行通行能力的分析，如，根据影像上的形状(线状、条带状)、色调、车辆数目多少，可分析出道路的类型、路面质量，交通堵塞情况及道路两旁的地形情况等，对部队的机动决策具有一定的参考价值。

②利用遥感信息进行军事地域类型分析。军事地域类型是指具有基本相同功能的地理实体集合在一定的地域上，形成面区位，且对战区作战起到基本一致的影响。分析军事地域类型对于战时的火力配系、兵力部署、战役方向选择有较大的实用价值。利用遥感图像，采用人机交互判读技术，根据不同地域类型所包含的具有代表意义的地理实体的影像特征来定性地判别出山地、平原、河流、丘陵、农业用地及植被覆盖情况，实质上是对影像上地形要素进行准确识别，利用自动识别技术及与DEM数据复合处理可得到有关地形的一些重要数据，如地面切割程度、坡度等数据，然后

利用分类方法作出定量分析，确定各地域类型的位置、面积大小、性质等。利用边缘提取技术和手动勾画界线方法，可制作出相应的军事地域类型区划图。

③利用遥感信息进行军事目标的分析。军事目标是指战区内重点保卫目标，也是敌人主要攻击点。它与军事交通网和军事地域类型构成了军事空间区位的主要内容。其中有些目标一旦遭到破坏会给部队的机动、供给造成困难，如战区内大中型城市，主要交通线上的小型城市、发电厂、油库等；有些目标被破坏后将给大中型城市造成重大损失，如石油加工厂、车站、自来水厂、电台、电视台、政府首脑机关等。军事目标的识别主要是在大比例尺航片上进行人机交互判读。遥感信息主要是来自地物对太阳辐射的反射作用，在影像上形成不同的灰度值和色调，因而，军事目标识别的依据主要是光谱特性，它能间接反映出各种地物的差异，可以通过影像的形状、大小、色调、阴影等特征来识别。另外，还需参考军事目标所处的地理位置、交通状况、绿化程度来识别。其中自动识别技术通过对基本的色调、灰度的统计分析，给出定量信息，然后利用人机交互判读技术可作出准确的判断，并在图上勾画和标注出来，为制作军事目标图提供位置信息和属性信息。

(4)利用 GIS 的空间数据进行遥感数据处理。利用 GIS 的空间数据可以提高遥感数据的分类精度。在利用遥感数据识别地物目标时，主要通过其光谱特性来间接反映出各种地物的差异。这样，常常发生“同物异谱”和“异物同谱”的现象，影响了分类的精度。而 GIS 中存贮的空间数据可以作为遥感数据处理的辅助信息源，以提高遥感图像分类的可信度。由于分类可信度的提高，又推动了 GIS 中数据快速更新的实现。

GIS 中的高程、坡度、坡向、土壤、植被、地质、土地利用等信息对遥感数据分类的影响很大，是遥感分类经常要用到的数据。

例如，森林覆盖的分类受高程和坡向的影响大。同样是森林，高度不同，生长的树种不同，阳坡、阴坡也不同。这种关系仅用遥感

资料是不能解决的，必须把高程和坡向的数据在地理位置上和遥感图像中的像元一一对应起来综合分析。具体方法为，将遥感数据与这两种数据配准，然后将高程和坡向分级。例如：小于 1000m、1000m～1500m、大于 1500m 三级高程，东南坡、西北坡、南坡等按方位角分为多种。根据实际情况将两个条件组合。例如，小于 1000m 的东南坡等。显然，不同高程与不同坡向的森林覆盖类别不同，如山上是针叶林为主，山下以阔叶林为主。这样就可解决由于光谱类似而造成的类别混淆，改善了分类效果。

总之，现有的基于光谱特征的多波段分类技术，仅利用了遥感数据所提供的最直接的信息，而未顾及到目标的空间分布及相互联系，而 GIS 的空间数据部分地隐含了这种空间关系，所以可在图像分类中作为潜在的知识源，以提高遥感数据分类的精确性和可靠性。

§11.2 全球定位系统(GPS)与 GIS 的结合

1. 全球定位系统(GPS)概述

(1)GPS 定位的基本原理。长期以来人类一直在为精确确定运动目标在地球上以及近地空间中的位置而进行不懈的努力。早期解决这一问题的主要方法是进行天文定位：测定太阳或其他天体的高度角和方位角并记录观测时间，以确定用户在该时刻的经纬度以及到某一目标的方位角。这种方法的定位精度通常不是很好，而且观测还要受气候条件的限制。

20 世纪初随着无线电技术的发展，各种无线电导航定位系统相继出现，如欧米伽(Omega)系统、罗兰 C(Loran－C)系统、台卡(Tacan)系统和微波着陆系统(MLS)等。无线电导航定位的基本原理与测量学中测定点位的方法十分相似，现以距离交会系统为例加以说明。设 A 和 B 为两个地面无线电信号发射台，其坐标已精确测定。用户 P 在某一时刻 t_i 采用无线电测距的方法分别测得

了从P点至A,B两点的距离SAP和SBP,那么我们只需以A点和B点为圆心,以SAP和SBP为半径作出两个定位圆,即可交出P点的平面位置。当然两圆相交一般会有两个交点,但用户根据其概略位置通常是不难加以判断和取舍的。况且为了提高定位的精度和可靠性,已知的无线电台站数量实际上往往不止两个,此时就不再存在两个解的问题了。根据用户至已知点的间距以及所需的精度,上述解算工作可以近似地在平面上进行,也可以在椭球面上进行。

与天文定位相比,无线电定位无论是在定位的速度还是在自动化程度方面都有了长足的进步,定位精度也有所改善,而且定位已基本不受气候条件的限制。然而地面无线电导航定位系统的作用距离(覆盖面)和定位精度之间会产生矛盾。低频率的无线电电波可以沿着地球表面传播,因而只要具有足够大的发射功率,信号就能传播到很远的地方。但长波信号的观测精度较差,而且信号的传播路径又难以准确确定,再加上信号需要在稠密的大气层中长距离传播,而用户又无法测定信号传播路径上的气象元素,因而难以进行准确的气象改正等,所以一些覆盖面很大的地面无线电导航定位系统的定位精度通常都较低。反之,使用高频率的信号时虽然可以获得较高的精度,但这些信号是沿着直线传播的,所以覆盖面很小。于是各个部门和单位为了满足各自的要求,相继建立了各种不同类型的地面无线电导航定位系统。

20世纪60年代,随着空间技术的发展,各种人造卫星相继升空。于是人们很自然地想到,如果把无线电信号发射机从地面站搬到卫星上,组成一个卫星导航定位系统,就能较好地解决覆盖面与定位精度之间的矛盾。全球定位系统就是一种采用距离交会法的卫星导航定位系统。如果用户P在某一时刻t_i用GPS接收机同时测得了从接收机至视场中的三颗GPS卫星(A、B、C)的距离S_{AP}、S_{BP}、S_{CP},而且该时刻这些卫星在空间的位置也是已知的话,那么我们就能用距离交会的方法求解P点的三维坐标(X_p,Y_p,Z_p)。其

观测方程为：

$$\begin{cases} S_{AP}=[(X_P-X_A)^2+(Y_P-Y_A)^2+(Z_P-Z_A)^2]^{1/2} \\ S_{BP}=[(X_P-X_B)^2+(Y_P-Y_B)^2+(Z_P-Z_B)^2]^{1/2} \\ S_{CP}=[(X_P-X_C)^2+(Y_P-Y_C)^2+(Z_P-Z_C)^2]^{1/2} \end{cases} \quad (11—2)$$

式中(X_A,Y_A,Z_A)，(X_B,Y_B,Z_B)和(X_C,Y_C,Z_C)分别为卫星 A,B,C 在时刻 t_i 的空间直角坐标。

由此可见 GPS 定位与通常的距离交会法从原理上讲并无差别，只是在一般的距离交会中，已知点 A、B、C 通常为固定点，其坐标一经确定后便能长期使用，而在 GPS 定位中三个已知点 A、B、C 是高速运动的卫星，其位置将随时间而迅速变化。此外在一般的距离交会中待定点至已知点的距离通常是用测距仪或钢尺丈量的，而在 GPS 定位中则采用了不同的方法。

GPS 全球定位系统是英文 Navigation Satellite Timing and Ranging/Global Positioning System 的字头缩写词 NAVSTAR/GPS 的简称。它的含义是：导航卫星测时和测距/全球定位系统。该系统是美军自 20 世纪 70 年代初开始研制的新一代卫星导航和定位系统，由空间部分、地面控制部分和用户接收机三部分组成。

全球定位系统的空间部分由 21 颗工作卫星及 3 颗投入使用的备用卫星组成。这些卫星分布在 6 个倾角为 55°的几乎为圆形的轨道上，每个轨道上有 4 颗卫星。卫星的平均高度为 20200km，运行周期为 12 个恒星时。GPS 卫星星座能提供全球性的覆盖。GPS 卫星使用 1.2GHz 和 1.6GKz 的高频波作为载波，信号沿直线传播，因而其定位的数学模型十分简单而又严格。此外由于信号只是从太空穿过大气层到达地面观测站，而不需在稠密的大气层中长距离传播(如地面无线电定位系统那样)，因而易于进行较为精确的对流层延迟改正。由于全球定位系统较好地解决了定位精度、定位速度、定位的可靠性及覆盖面等因素之间的矛盾，因面有可能逐步取代相互交叉重叠的地面无线电导航定位系统，而成为一种通用的导航定位系统。

(2)GPS 的测距方法。

——概述。我们知道测距通常可采用几何测距和物理测距这两种方法。几何测距是通过将待测距离与标准长度(如铟钢线尺、钢卷尺等)进行几何对比来实现的。这种方法显然不能用于测定用户与 GPS 卫星间的距离。物理测距则是通过测定测距信号的传播时间来间接测定距离的。在物理测距中信号的传播速度通常被认为是已知的(信号传播过程中介质对传播速度的影响通常是通过专项改正来加以消除的)。物理测距又可分为下列两种方式：

主动式测距。用户发出测距信号，该信号传播至测线另一端时被反射回来。用户用仪器及时钟(直接或间接地)测定测距信号往返传播的时间 Δt 后即可用下式求出距离 S：

$$S=\frac{1}{2}\Delta t\cdot V+\sum\delta_i \qquad (11-3)$$

式中 V 为信号的传播速度，δi 为各项改正数。用激光测距仪来测定至卫星的距离，用雷达测高仪来测定卫星至瞬时海面的高度(卫星测高)，用声纳仪来测定水深以及地面测量时所用的各类测距仪都采用这一测距方式。采用主动式测距方式时，信号离开仪器及重新返回仪器的时间都是由用户时钟来测定的，因此用户通常只需采用短期稳定度良好的石英钟就能获得精确结果。

主动式测距的缺点是用户需发出测距信号，因此在战争的环境中难以隐蔽自己，此外要同时测定至几个目标的距离也较为困难，而且用户还必须同时配备发射设备和接收设备。

被动式测距。位于测线另一端的无线电信号发射机(置于卫星或地面台站中)按照预定的方式在预定的时刻发射测距信号，用户则用自己的钟来测定信号到达的时间，求出信号的单程传播时间 Δt 后用下式计算距离：

$$S=\Delta t\cdot V+\sum\delta_i \qquad (11-4)$$

采用被动式测距方式时信号是按照无线电发射台中的时钟来发射的，而信号到达时间却是用用户时钟来测定的。因此按上式求

得的距离 S 中必定会包含这两台时钟不同步的影响，所以用这种方式测得的距离被称为伪距。采用被动式测距方式时用户只需接收信号，易于隐蔽自己。因而这种方式乐于被军事用户所采用。然而无线电信号是以每秒 30 万公里的速度传播的，所以采有这种方式测距时必然会对两台时钟的同步误差提出极高的要求。例如如果我们要求同步误差对测距的影响不大于 30cm，那么这两台时钟就应该同步至 1ns（10^{-9}秒）以内。假设这两台时钟每 3 小时可精确比对一次且不计比对误差的话，那么我们就需对这两台时钟提出如下要求：在 3 小时内它们的钟差不得超过 1ns。因而每台时钟的稳定度至少应优于 10^{-13}。只有性能极好的高质量原子钟才能达到这种精度。每个 GPS 工作卫星上都配备了 4 台原子钟，然而在数以万计的 GPS 接收机上都配备原子钟是不现实的，因为这将使接收机的体积重量和价格大幅度增加。为解决这一问题，在用户接收机中一般仍配备石英钟，但需把观测瞬间用户时钟的钟差 V_T 也当作一个未知参数，通过平差计算来求得它。这样在 GPS 定位中就至少有了 4 个待定参数，因此在 GPS 定位中用户通常同时测定 4 个 GPS 卫星的距离才有可能解得这些参数。其数学模型为：

$$\begin{cases} S_{AP}=[(X_P-X_A)^2+(Y_P-Y_A)^2+(Z_P-Z_A)^2]^{1/2}+(V_{tA}-V_T)\cdot C \\ S_{BP}=[(X_P-X_B)^2+(Y_P-Y_B)^2+(Z_P-Z_B)^2]^{1/2}+(V_{tB}-V_T)\cdot C \\ S_{CP}=[(X_P-X_C)^2+(Y_P-Y_C)^2+(Z_P-Z_C)^2]^{1/2}+(V_{tC}-V_T)\cdot C \\ S_{DP}=[(X_P-X_D)^2+(Y_P-Y_D)^2+(Z_P-Z_D)^2]^{1/2}+(V_{tD}-V_T)\cdot C \end{cases} \tag{11—5}$$

式中 V_{tA}、V_{tB}、V_{tC}、V_{tD}分别为观测瞬间 4 个卫星的卫星钟的钟差，可以从卫星星历中求得。S_{AP}、S_{BP}、S_{CP}、S_{DP}中已加入了电离层延迟改正、对流层延迟改正等各种改正项。GPS 卫星星座可以保证任一时刻地球上任何地方的用户至少能同时看到 4 颗 GPS 卫星。

——测距码。全球定位系统是利用测距码来测定距离的。测距码是一种周期性的按照某种特定规律编排起来的二进制代码。

这是一种从某一特定结构的线性反馈移位寄存器中产生并经过移位、复合、截短等处理后形成的一种伪随机噪声码。如果用户接收机也能产生结构完全相同的测距码(称为复制码或本地码),那么就能用它来测定从接收机至卫星间的距离。其具体做法如下:

卫星和用户接收机在各自时钟的控制下产生出结构相同的测距码和本地码。卫星发出的测距码经过 Δt 的传播后到达接收机被接收机接收,并与通过时延器延迟后的本地码进行比对。不断调整延迟时间,使来自卫星的测距码与延迟后的本地码“对齐”,此时本地码的延迟时间 τ 即为卫星信号的传播时间 Δt。而伪距测量实质上就是测定信号的传播时间。在 GPS 测量中测距码和本地码是否已对齐是根据这两组信号的相关系数 R(t)的大小来判断的。t 时刻的相关系数 R(t)的定义如下:

$$R(t) = \frac{1}{T}\int_T U(t - \Delta t) - U'(t - \tau)dt \qquad (11—6)$$

式中 U 为卫星发出的测距码,U′为接收机发出的本地码。它们都是用幅度为 1 的正负矩形波来表示的,正波形(+1)代表二进制码“0”,负波形(−1)代表二进制码“1”。也就是说将接收到的来自卫星的测距码 $U(t-\Delta t)$和延迟后的本地码 $U'(t-\tau)$相乘,并在时间段 T 内进行积分,其积分平均值即为相关系数 R(t)。由于测距码和本地码的结构相同,因此当 $\tau=\Delta t$ 时相关系数 R(t)将达到最大值 1,当 $\tau\neq\Delta t$ 时相关系数 $R(t)<1$。伪随机噪声具有良好的相关特性,当时延 τ 略偏离 Δt 时相关系数 R(t)便会迅速减小。采用伪随机噪声码作为测距码有助于精确测定距离。GPS 接收机能自动调整延迟时间使相关系数始终保持最大(相关系数最大时其理论值应为“1”),然而由于时钟误差的影响以及信号传播过程中的畸变等原因,所以在进行相关处理时两种码并不完全相同,相关系数也无法达到 1,这样用户就能很方便地求得任一时刻从接收机至卫星的伪距。

GPS 卫星的测距码分为两种,一种是粗码/捕获码(C/A 码);

另一种是精码(P 码或 Y 码)。粗码的码元长度(每个码所持续的时间与光速 C 的乘积)为 293m,精码的码元长度为 29.3m。相关处理的精度大约为码元长度的 1/100 左右。因此用精码来测距可以获得更高的精度。全球定位系统投入正式工作后,美国对精码进行加密,控制其使用范围。

利用测距码进行测距的优点如下:

①GPS 卫星的发射功率约为 20W,卫星离用户的距离却超过 2 万公里。而一般无线电台的发射功率约为 1000W,距用户的距离通常只有几十至几百公里。因此 GPS 卫星的信号是非常微弱的,被深深掩埋在各种干扰信号中。其信号强度通常要比噪声的强度低 3~4 个数量级。用测距码来进行测距并将其调制在载波上发射出来(即采用所谓的扩频技术)后,我们才有可能从噪声的汪洋大海中将 GPS 信号提出来。

②采用相关法测距时,我们并不是根据某一码(如脉冲法那样)来测定距离的,而是在相关系数最大,也就是说从总体上讲整个积分区间内所有的码都对得最齐的情况下测定距离的。从某种意义上讲,相关法所测定的距离可以看作是用积分区间中的每个码所测定距离的平均值,所以具有较高的精度。

③系统的研制管理机构可以通过公开码结构或对码结构进行保密的办法来对用户使用该系统进行控制。例如公开 C/A 码的结构,允许全球所有用户利用 C/A 码来进行测距和定位;对 Y 码进行保密,只允许美国及其盟国的军方用户以及经美国政府批准的用户使用 Y 码来进行测距和定位。

④便于对不同 GPS 卫星的信号进行识别和处理。GPS 接收机只需用一个全向型的天线就可接收视场中所有 GPS 卫星的信号。由于每个 GPS 卫星所用的测距码的结构都互不相同,而且均相互正交(不同卫星的测距码相乘并进行积分后,无论这两个码的相对位置如何,其相关系数总是趋于零),因此凡是和本地码结构不同的卫星信号都将被自动清除,只有与本地码结构相同的卫星信号

才能与之进行相关处理。因此如果视场中有6颗卫星，那么接收机只要拿出6个通道来让每个通道观测一个卫星信号(即每个通道中都产生与它所要观测的卫星的测距码结构相同的本地码)，就能同时测得至6个卫星的距离，彼此互不干扰。

——载波相位测量。利用测距码进行伪距测量是全球定位系统的基本测距方法，然而由于测距码的码元长度较大，对于一些高精度应用来讲其测距精度还显得过低，无法满足需要。加之美国政府还将对P码进行加密以控制其使用范围，使上述矛盾变得更为突出。载波相位测量技术就是在这种背景下产生的。在无线电通讯中载波只是一种运载信号的工具。而在全球定位系统中载波不仅被用来运载测距码和卫星电文，而且本身又被当作一种测距信号。为了消除电离层延迟误差，GPS卫星同时使用了两种频率不同的载波：频率为1575.42MHz的L1载波和频率为1227.60MHz的L2载波。L1载波的波长约为19.0cm，L2载波的波长约为24.4cm。载波相位测量的精度可以达到毫米级至亚毫米级。然而载波的波长只是一些无任何标记的余弦波，因而进行载波相位测量时会产生整周模糊度的问题。也就是说，对某一卫星进行首次载波相位测量时我们能测定的仅仅是不足一整周的部分，而此时从接收机至卫星间究竟包含了多少个整周是无法测定的。在随后进行的各次载波相位测量中，虽可记录下整周数的变化量(称为整周计数)，但由于各种因素的影响，记录下来的整周计数常常会产生跳变，使后面的整周计数都产生偏差。因此处理载波相位测量的观测值之前，通常需花费不少时间和精力首先对整周跳变进行探测——找出何处发生了整周跳变，然后进行修复——确定跳变的周数并对随后的整周计数逐一进行修正。由于载波相位测量中存在着整周跳变和整周模糊度的问题，因此数据处理通常要比用测距码测距时复杂得多，困难得多。这是为了获得高精度所需付出的代价。用户可以根据不同的精度要求采用不同的测距方法。

在全球定位系统的设计阶段，一些大地测量学家曾建议在

GPS 卫星中设置 10 种不同的频率。这样用户就能用 10 种不同波长的尺子来测量距离，每种波长依次相差一个数量级。采用上述方法后能精确测定从接收机至卫星的距离，而不会产生整周模糊度和整周跳变的问题。遗憾的是上述建议并未被采纳。因为全球定位系统的主要目的并不是为了满足精密大地测量的需要，而是为了满足军事用户导航定位的需要。此外对用户使用该系统的能力进行控制也是系统设计中必须加以考虑的重要因素。

(3)GPS 的卫星星历。

——地面控制系统。如前所述，进行 GPS 定位的一个先决条件是用户必须知道观测瞬间 GPS 卫星在空间的位置。由于不同的用户将在不同的时间对不同的卫星进行观测，所以我们实际上是要求知道所有的 GPS 卫星在任一时刻的位置。上述要求是在全球定位系统的地面控制部分的支持下得以满足的。全球定位系统的地面控制部分是由一个主控站(位于美国科罗拉多州)，三个注入站(分别位于太平洋的卡瓦加兰岛，印度洋的狄哥·伽西亚和大西洋的阿松森岛上)，5 个监测站(除位于上述四地外，再加上夏威夷群岛)，以及通讯辅助系统组成的。

监测站是无人值守的 GPS 卫星跟踪站，其站坐标已精确测定。每个监测站内均配备有双频 GPS 接收机，气象传感器，原子钟和微机。以原子钟作为频率标准的 GPS 接收机在微机控制下对视场中的所有 GPS 卫星进行伪距测量。为了对观测值进行对流层延迟改正，气象元素传感器还可自动采集当地的温度、气压和相对湿度。在微机控制下监测站能对伪距观测值进行各项改正，并对观测资料进行编辑、平滑和压缩，并通过通讯系统将资料送往主控站。主控站对各监测站送来的观测资料进行处理，以计算每个卫星的运行轨道以及卫星钟的改正参数，并外推出未来 26 小时的卫星轨道和卫星钟参数。然后按规定格式编制成卫星导航电文送往注入站。再通过各注入站将它们送往每个卫星，寄存在卫星的内存中。卫星则将导航电文调制在测距码的载波上按时播发给用户。用户

用相关法进行伪距测量时,可同时获得卫星星历,经简单计算后即可求得该时刻卫星在空间的位置。

——广播星历和精密星历。由于上述卫星星历是通过GPS卫星"广播"给用户的,所以称为广播星历。广播星历的精度较低,这是由于:

①广播星历是一种外推星历或者说预估星历,由于我们对影响卫星运动的各种因素的变化规律了解得还不是十分透彻,预估的情况往往会和实际情况有所出入,从而影响星历的精度。

②美国政府实施SA技术,有意在星历中加入误差,使广播星历的精度大幅度下降。

鉴于卫星星历在GPS定位中的重要作用,而广播星历的精度又低这一实际情况,目前有不少国家和部门正在考虑(或已经)建立自己的GPS卫星跟踪网,开展独立定轨工作以获得较为精确的GPS卫星星历。这些星历被称为精密星历,它们主要用于GPS观测资料的事后处理。

(4)美国政府的GPS政策。由于全球定位系统在军事上具有重要的作用,因此美国政府已决定采用SA技术(Selective Availability)和AS技术(Anti-Spoofing)把未经美国政府特许的广大用户的实时定位精度降低到它所允许的水平:±100m(2drms),以免美国的国家利益受到损害。

——SA技术。SA技术主要包括下列两个内容:

①有意识地在广播星历中加入误差,人为降低星历的精度。

②有意识地在卫星钟频信号中加入误差,使钟频产生快速变化(即所谓钟频抖动),据报道,这种钟频变化可达±2Hz(标准频率为10.25MHz),变化周期约为10分钟左右。

采取第①项措施是为了降低GPS定位中已知点的坐标的精度。采取第②项措施则是为了降低测距的精度。当然在实施SA技术时需保证美国及其盟国的军方用户以及经美国政府批准的少数用户定位精度不受影响,也就是必须让这些内部用户知道究竟加

入了多少误差以便加以修正。为此主控站又将上述数据编入只调制在保密的精码上的卫星星历中，供上述内部用户使用。

——AS 技术。所谓 AS 技术指的是对精码的码结构进行保密，以防敌对方进行电子干扰和电子欺骗的技术。其具体措施是将 P 码与保密的 W 码相加以形成 Y 码。Y 码的码结构是严格保密的，只有美国及盟国的军方用户及经美国政府批准的内部用户才能使用它。

——影响及对策。综上所述，美国政府的 GPS 政策从实质上讲就是想把用户分为一般用户和内部用户两类，做到内外有别。内部用户可以使用精码以及未降低精度的卫星星历来进行定位，以获得尽可能高的定位精度，这就是所谓的精密定位服务(PPS)。而其余的用户则只能用 C/A 码以及已被人为降低了精度的卫星星历来进行定位，获得美国政府所允许的定位精度，这就是所谓的标准定位服务(SPS)。目前 PPS 的定位精度为 16m(SEP)，SPS 的定位精度为 100m(2drms)。

SA 技术的实施也会对进行事后处理的测量用户产生一些消极影响。建立 GPS 卫星跟踪网，独立确定 GPS 卫星的精确轨道可以从根本上消除上述消极影响。我国目前也正在着手建立自己的 GPS 卫星跟踪网。由于种种原因，2000 年 5 月 1 日，美国总统克林顿签署命令，宣布解除对 GPS 卫星信号的 SA 限制，GPS 用户因而可获得比原来高 10 倍的定位精度。

2. GPS 与 GIS 的结合

全球卫星定位系统(GPS)作为一种新型的定位数据的采集和更新手段，具有高精度、高效益、全天候、低成本、高灵活性、实时性等特有的优势，因而在 GIS 中具有重要的应用价值。

GPS 在 GIS 中的应用常常分为两种情况：

一是直接用 GPS 技术对 GIS 的空间数据作实时更新和采集；

二是把 GPS 接收机的实时差分定位技术与 GIS 的电子地图相结合，组合成各种电子导航系统。

另外,GPS 还可以为 GIS 中空间数据的采集提供辅助的定位数据,可大大提高成果数据的精度和应用范围,例如在航空摄影测量和遥感中应用 GPS 技术。

下面把 GPS 在 GIS 的有关应用举例说明。

(1)GIS 空间基础数据的采集。这种数据一般包括应用 GPS 技术所建立的大地测量控制网和水准模型等数据。这些数据所体现的大地模型是地理信息系统所有空间数据赖以存在的空间基准,它一经建立就会保持相当长时间的稳定,可保持几十年甚至上百年。目前,应用 GPS 系统的静态或快速静态观测模式,可建立各种等级的测量控制网。由于 GPS 技术的应用,使得各等级的测量控制网的布设在精度、速度和成本等方面得到了极大的改善。

(2)地形数据的局部修测。地形数据是 GIS 数据的基础部分,这些数据在总体上是比较稳定的,整体数据的更新时间比较长,但局部的修测和补充往往比较频繁,否则数据的现势性和应用价值就会降低。利用 GPS 来进行这方面的工作,具有明显的优越性。传统的地形测量方法,必须遵循先逐级布设控制,然后再进行碎部测量的操作程序。并且,要按照所布设的图根点逐一设站,每站测量的地形范围都是有限的,所以工作效率和成图精度都受到了很大的限制。而利用 GPS 的 RTK 定位技术进行野外作业时,流动站(工作站)与参考站之间的距离可达 10～20km,所以,流动站可在参考站周围 10～20km 的范围内机动灵活地采集数据,不受视距长度和地形条件的限制,大大提高了工作效率。RTK 定位技术本身的精度可达到厘米级,由于没有其他的中间环节,所以成图的精度也是很高的,碎部点点位数据信息的采集精度将优于 0.1m。

(3)公路数据的采集与更新。在地理信息系统中,有关公路尤其是高等级公路的数据包括公路横断面和纵断面以及中心线等数据。GPS 技术非常适合于公路信息的采集工作。横断面数据的采集以利用 GPS 的准动态观测模式为宜。横断面测量一般每隔一定的间隔采集一个数据面。公路纵断面数据信息的采集,可利用动态

GPS 定位技术。采集公路纵断面数据信息时，一般沿公路的左边线、右边线和中心线连续地采集断面点的坐标信息。如果应用传统的采集方法，需沿着公路断面线每隔 20m 采测一个点，与之相比，动态 GPS 定位技术的一个明显的优势是它能以更高的采样密度采集断面点的平面坐标和高程信息，从而可以获得更为逼真的公路断面图。加拿大的测量工作者在一次试验性的公路断面测量工程中，通过精度的检核说明，在 99%的置信水平下，应用动态 GPS 定位技术采集断面点的点位信息在水平和垂直两个方向的精度都优于±5cm，高程精度 68%优于±16mm，95%优于±32mm。由此可见其精度足以满足实际要求。

由于大量的公路建设和改造，有许多新公路的数据信息需要进入地理信息系统，因此有时要进行大面积的公路网数据信息的普查工作。这项工作的主要任务是采集公路中心线的位置。这种数据信息采集工作的精度要求一般并不很高，如果是与 1:5 万或 1:10 万等中小比例尺地形数据相结合，能达到米级的采样精度就能满足要求，这时可利用 GPS 伪距动态差分定位技术。观测时，参考站可放在一个比较方便的地方，流动站可安置于汽车的车顶，沿着公路的中心线行驶，这样每天可测量上百公里的线路。如果是为 1:1 万以上的中、大比例尺地形数据相结合，则要求成果精度优于 1.0m，为此，可应用载波相位差分动态测量技术，相位差分动态测量要求流动站和参考站之间的距离小于 20km，这时为了提高工作效率，可沿路每隔 30km 左右连续布设几个参考站，只设置一个车载流动站，这样，一次就能连续测量近百公里左右的线路。所以，利用车载 GPS 动态测量，能以极快的速度完成这类线路普查生产任务。

(4)边界数据的采集与更新。在各种境界测量中，GPS 定位技术同样能显示其优越性。如在县界测量中，边界点有很多都位于很偏僻地区，距离已知的控制点很远，但是，边界点的分布都比较密集呈线状分布，此时根据边长的长短情况，可灵活运用 GPS 的静

态和快速静态观测模式。一般先沿着边界的走向，利用静态观测模式每 20km 左右布设一个控制点，然后，以这些控制点为参考站，利用快速静态模式，对其周围方圆 20km 范围内的边界点逐一测量，可极大地提高边界测量的精度和工作效率。

(5)周期性数据的采集。周期性数据主要是指在地震和变形监测等领域采集的周期性数据，这些数据的特点是更新速度快而且新旧数据间要相互比较。用于地震监测和预报的控制网中，因为基线都比较长，所以在实际测量中采用静态观测模式，在这个领域里，GPS 技术有着其他技术无法替代的优势，它能极大地提高地震监测的精度、缩短检测周期，在关键时期可以将观测周期缩短到一天甚至几个小时，从而提高了监测可靠性。

在普通的形变监测中，因为基线长度比较短，所以，一般采用快速静态观测模式。普通的快速静态测量的精度为±5～10mm，这还不能满足形变监测的要求，为了满足形变监测±1～2mm 的测量精度，必须采取一些特殊的措施，如强制归心、传感器定向和选择多卫星状态等等。由于 GPS 测量具有全天候的特点，使其在某些形变监测中发挥了重要的作用。例如，在大型水坝的形变监测中，一旦汛期来临，监测的频率必然提高，但是，由于汛期的雨水不断，传统的光电测量难以进行，如果采用 GPS 技术，只要不是雷雨，采取一些简单的防雨措施就可以全天 24 小时频繁工作。

(6)GPS 为 GIS 提供实时定位信息。由于 GPS 可以提供实时的定位信息，因此，当把 GPS 与 GIS 连接起来后，用户可以很快地在 GIS 的电子地图上找到自己的位置。因此在旅游、探险、航行、军事等领域均有广泛的应用，如用于车、船的定位和自动驾驶等。GPS 与 GIS 的结合方式如下：

· GPS 单机定位＋栅格电子地图。该系统可实时地显示移动物体(如车、船、飞机等)的所在位置，从而进行辅助导航。优点是价格便宜，缺点是精度不高，功能简单。

· GPS 单机定位＋矢量电子地图。该系统可以根据目标位置

(工作时输入)和车船现在的位置(由 GPS 测定),自动计算和显示最佳路径,还可以用多媒体方式向用户提示。但矢量数据库的建立需花费较大成本,同时 GPS 的测量误差应设法加以补偿和改正。

·GPS 差分定位+矢量/栅格电子地图。该系统通过固定站与移动车船之间两台 GPS,并使用伪距差分技术,可使定位精度达到 1~3m。这种情况需有通讯联系,可以是单向的或双向的。即 GIS 可放在固定站上,构成车船现状监视系统;也可以放在车船上,构成自导航系统;双方均有 GIS 加通讯,则可构成交通指挥、导航、监测网络。

第十二章 网络GIS

§12.1 引 言

GIS与地图有着天然的联系。使用图形进行信息表达和信息交流几乎和人类一样的悠久，地图作为地理信息的载体已经为人类服务了几千年，其中的一点优势在于，地图的使用几乎不存在语言的障碍。今天，随着计算机技术的发展进入以因特网(Internet)和万维网(WWW，World Wide Web)为代表的网络时代，地图的使用甚至解除了“地理位置的束缚”，进入了拥有万维网浏览器的千家万户。GIS通过万维网扩展，已经真正成为一种大众使用的工具。从万维网的任意一个节点，因特网用户可以浏览万维网站点中的空间数据，制作专题图，进行各种空间查询和分析。

随着计算机网络的发展和普及，用户的需求已不再限于桌面上的地图应用和地理分析，新的技术和需求使计算机网络体系下的GIS技术与应用发生了重大的变化，地理信息资源的分布和共享成为网络时代GIS技术所要解决的基本矛盾。

1. GIS技术与应用的变化

(1)操作平台和计算环境。GIS的操作平台和计算环境的演化大致经历了三个阶段：集中式处理系统阶段、客户/服务器(C/S，Client/Server)系统阶段和分布式计算系统阶段。进入20世纪90年代以后，计算环境更是朝着通过网络把许多专用GIS和地理信息资源有机集成为特征的体系结构进化。其实，自然界也在验证着类似的发展趋势。例如，单细胞有机体进化成复杂的多细胞生物体(复杂的系统往往具有更强能力)。细胞逐渐特化(功能专一)，因为

小的专用部分往往比大的通用部分更容易成功地进化而实现理想的功能。在一个计算环境中，细胞即一个独立的系统。通过连接网络上所有系统节点，可以把各个专用的系统组合成一个集成的计算环境。

（2）软件技术。突出体现为 GIS 软件的组件化。也就是说，大的 GIS 软件被分解为许多具有不同功能组合、可互操作和自我管理的功能组件；符合一定技术标准的 GIS 组件能灵活方便地集成出具有特定用户模式的 GIS。目前支持这种分布式部件对象模型的技术标准主要有三种：OMG 的 COBRA（Comon Object Request Broker Architecture）、Microsoft 的 DCOM（Distributed Component Object Model）和 SUN 的 Java。

（3）数据组织。传统的 GIS 以统一、完整的空间数据库系统存储和管理海量空间数据，而互联网环境下则是多种不同类型的地理数据分布在成千上万的网络站点上，形成多层嵌套、无限扩展的地理空间。因此，分布式异型地理数据的组织、管理、查询、分析、显示、发布，是网络 GIS 所要解决的一系列重大问题。

（4）用户。在互联网环境下，GIS 将面对不同应用水平、不同使用目的、不同文化背景的各种网络用户群，因此需要更加客户化的应用模式和更加严密的安全机制。互联网时代的 GIS 应用更多的是一种综合信息服务的整体解决方案，而不仅仅是软件产品。

2. 计算机网络环境下 GIS 要解决的基本矛盾

（1）地理信息的分布式本质。地理信息资源是一种重要的分布式信息资源。地理信息本质上是分布式的，而不是集中式的。从空间角度看，地理数据是分布在空间区域中的，可以根据行政、自然地理区域等等来组织地理数据。采集、管理和维护这些不同区域的数据并没有因果从属关系，而是相对独立和平行的。例如：广州和上海的城区数据是由这两地区分别管理的。高效的系统应该是当地的数据变动由当地处理，但各部门都可以在网络上各个节点调用其他区域的数据。从特性的角度来看，地理信息可以以其自然性

质和人文性质分类组合。每个机构有许多部门和分支机构，这些部门根据各自的分工有责任收集、保存、更新属于自己范围内的数据。例如，水文部门、城市规划部门、环境保护部门、公共事业部门都在他们各自权力范围内从工作中采集和管理各自的数据，但又需要调用其他部门的资料帮助他们的决策工作。从应用角度看，机构内不同部门的用户可以使用 GIS 来完成他们日常的工作，而外围用户则需要通过多用户的方式来使用 GIS。这些部门对不同地区和类型的地理信息具有不同的需求。地理数据一个主要的特征是分布而不是集中，因此分布式的计算机网络系统最适合于处理地理数据。这是当前 GIS 发展的一个趋势。

(2)地理信息的共享是 GIS 在企业、政府和社会中应用成功的关键。由于地理信息的数据模型、数据结构以及处理方法的复杂，加上处理方法的独立和产品种类的差异，使地理信息的共享和应用受到极大的制约。在企业、政府或整个社会环境中，一个好的 GIS 系统应该可以提供强有力的计算机资源、技术资料和数据的共享功能。基于以下三个原因，GIS 一定要在一个企业、政府的多个部门甚至整个社会共享：

· 单独重复地收集和管理数据会造成系统投资和资源的浪费。

· 由于不可能长期保持多个独立数据库的统一维护和更新，应用不同的数据会在决策中造成不一致。这种不一致包括不同的坐标系统，比例尺，数据来源，数据更新时间，数据的分类和结构等。要在这种数据上做出决策将会是低效和混乱的。

· 复杂决策过程需要用到其他不同层次的数据，这就决定了数据共享的关键性。一个实用的数据共享形式是各部门有充分的权利管理自己的数据，而又可以得到其他部门的数据来辅助自己的决策。通常为防止数据被破坏，一部门应该只有权限修改管理和更新自己管理的数据，而没有权利去随意修改其他部门的数据。网络上的企业 GIS 系统应该提供方便而可靠的技术手段去支持这

一工作方式。

除数据外，计算机的软硬件也应该在企业环境中共享。这包括CPU 的能力，储存空间，周边设备，软件功能和应用模型。实际上，许多 GIS 功能可以在网络上从一个节点流动到另一节点。例如软件包浮动许可证这一概念就很适合企业化的工作方式，因为用户可以利用其他部门暂时闲散的软硬件。共享数据和服务将会让企业内各部门，而不是一个部门受益。

地理信息的分布和共享之间的矛盾给 GIS 技术和 GIS 产业带来巨大的机遇和挑战。一方面，把网络技术引入 GIS，开拓了地理信息资源利用的新领域，有利于地理信息的产业化和社会化。另一方面，开发和研究人员将重新审视 GIS 原有的体系结构、数据组织、软件组织以及应用模式，并制定新的技术方案，开发新的产品，给 GIS 的大家族增添新的成员——网络 GIS。

本章我们将简单介绍计算机网络的基本概念，然后介绍当前网络 GIS 的实现技术以及各主要 GIS 厂商的网络 GIS 解决方案。

§12.2 计算机网络的基本概念

过去的 300 年中，每个世纪都有一种主流技术。18 世纪伴随着工业革命而来的是伟大的机械时代；19 世纪则是蒸汽机时代；而在 20 世纪，关键技术是信息收集、处理和发布。我们已经看到了世界范围的电话网、电视机的发明，计算机工业的诞生及其史无前例的迅速发展，通讯卫星的发射，以及其他种种成就。

由于技术的飞跃发展，这些领域正在迅速的融合。信息收集、传送、存储和处理之间的差别在迅速的消失。在广阔的地理位置上分布的成千上万的办公室机构，互相可以了解千万里之外遥远地点的最新情况。

计算机和通信的结合，对计算机系统的组织方式产生了深远的影响。一间大屋子，装备一台大型计算机，用户带着他们的工作

去上机处理的“计算机中心”概念现在已经完全过时了。单台计算机为一个企业或机构中所有的计算机需求服务这一概念，很快被大量分散但又互联的计算机来共同完成的模式所取代。这种以共享资源为目的，通过数据通信线路将多台计算机互连而成的系统被称为计算机网络。

当个人计算机网络提供了比主机高很多的性能价格比时，人们对计算机网络感兴趣是无庸置疑的。使用网络的第一个目的是资源共享和通信，网络上的用户，无论身处何地，也无论资源的物理位置在哪，都能使用网络中程序、设备，尤其是数据。通过网络，不同地方的人还可以一起写工作文档和信件，而不必花几天的时间来等待这种人际间的沟通。第二个目的是依靠可替代的资源来提供高可靠性。多处理机的出现，文件可以在多台计算机上留有副本，意味着如果其中的一台机器出了故障，系统仍能继续运行。还有一个目的是节约经费和可扩充性。主机比个人计算机大概快十倍，但价格却在千倍以上。采用主机方式，当系统能力达到极限时需要更换更强大的主机，而使用网络，只需增加更多的处理器，就能逐步改善系统的性能。

从 20 世纪 90 年代开始计算机网络开始为居家的个人用户提供服务。这些激动人心的服务主要有：

(1)访问远程信息。这包括访问财务部门，查询自己的银行户头；访问商场浏览成千上万的联机货物清单；阅读电子报纸和在线图书馆；访问信息系统，获得有关艺术、商业、餐饮、政府、健康、娱乐、地理、历史、科学、体育、旅游等方面的信息。

(2)个人间通信。这可能是 21 世纪替代 19 世纪发明的电话的手段。电子邮件(email，electronic mail)已被广泛使用，并且能将声音、文本、图像一起传送。实时的电子邮件使远程用户可以无延迟地通信，并且可以互相看到或听到对方。这种技术可以用来召开虚拟会议，即视频会议(video conference)，虚拟会议可被用于远程学校、远方专家会诊以及新闻沙龙。

(3)交互式娱乐。最吸引人的是视频点播(video on demand)和游戏,这将是一个新兴的、基于计算机网络的巨大工业。

本节将介绍与网络GIS技术有关的一些计算机网络的基本概念。在介绍网络GIS技术中将涉及这些概念和技术。如果希望详细了解计算机网络,请参阅有关计算机网络方面的书籍。

1. 局域网、城域网、广域网和互联网

按照连接距离的长短,计算机网络可分为局域网(LAN,local area network)、城域网(MAN,metropolitan area network)和广域网(WAN,wide area network)。

(1)局域网。局域网的覆盖范围比较小,通常是同一建筑、同一大学或方圆几公里区域内的专用网络。局域网通常使用这样一种传输技术,即用一条电缆连接所有的机器。传输速度可以达到数百兆比特/秒(Mb/s),传输延迟低,并且出错率低。局域网大多采用广播式网络,它仅有一条信道,由网络上所有机器共享,它可以有多种拓扑结构。图12—1给出了最常用的两种:总线型和环型。

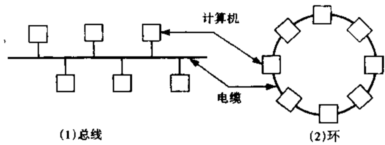


图12-1

(2)城域网。城域网通常覆盖一组临近的公司办公室或一个城市,通常使用与LAN相似的技术,但往往使用两条单向电缆,所有的计算机都连在上面。

(3)广域网。广域网是一种跨越大区域的网络,通常包含一个国家或省。它由主机(host)和通信子网(communication subnet)的

集合组成。子网是把数据从源主机传送到目的主机的通信线路和路由器(router)的集合。

当网络之间需要通信时,就需要连接不同的、往往互不兼容的网络,通常使用被称为网关(gateway)的机器来连接,并提供硬件和软件的转换,互联的网络集合被称为互联网。常见的互联网就是通过 WAN 连接起来的 LAN 集合。

(4)互联网。两个或更多的网络的连接被成为互联网(internetwork)。

2. 因特网

因特网是一个连接世界各地不同计算机网络系统的全球性网络系统。是一种特定的世界范围的互联网。它所基于的通讯协议是传输控制协议(TCP,Transmission Control Protocol)和互联网协议(IP,Internet Protocol)。因特网可以把来自全世界各个国家、地区、各个不同机构、公司、政府部门和院校的信息迅速传播到地球的每个角落。随着因特网的应用——万维网的出现,它已经成为目前世界上规模最大、增长最快的计算机网络系统,被称为"信息高速公路"的雏形,影响着人们的工作和生活方式。

(1)因特网的 4 种主要应用。电子邮件(Email)。现在几乎各种计算机上都有电子邮件程序。很多人认为这是和外界联系的基本方式。比电话和缓慢的邮件好得多。

新闻(news)。新闻组是特殊的论坛。现在已有上千上万个新闻组,关于技术和非技术专题的,包括计算机、科学、娱乐和政治。每个新闻组都有自己的规则、风格和习惯。

远程登录。通过 Telnet,Rlogin 或其他程序,因特网上的任何用户都可以登录到他们有帐号的机器上,使用其他机器的软件和资源进行工作。

文件传输。通过 FTP 程序,用户可以从因特网上的一台机器向另一台机器拷贝文件。可以用这种方式获取大量的文章、数据和其他信息。

(2)传输控制协议和互联网协议（TCP/IP)。因特网按层的方式来组织，一台机器的第 n 层与另一台机器的第 n 层(对等实体)进行对话，通话的规则就是第 n 层协议(protocol)。实际上，数据不是在对等实体之间传输，而是每一层都把数据和控制信息交给它的下一层，直到最下层。第一层下是物理介质，它进行实际的通信。在图 12—2 中，虚线表示虚拟通信，实线表示物理通信。每一层通过接口为它的上一层提供服务，实现服务的细节对上一层是屏蔽的。

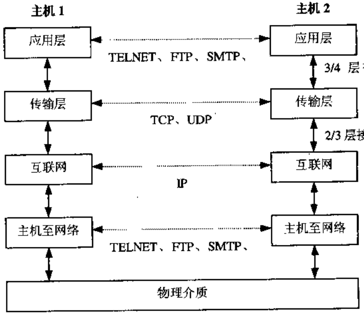


图 12－2 因特网的组织方式

在卫星和无线网络出现以后，TCP/IP 以前的协议在和它们互联时出现了问题，这时需要一种能无缝连接多个网络的网络体系结构，这个体系结构在它的两个主要协议出现以后，被称为 TCP/IP 参考模型(TCP/IP reference model)，如图 12—2 所示。

· 互联网层(internet layer)。它是整个体系结构中的关键部

分，定义了网络中数据的传输单元，正式的分组(packet)格式和协议，即IP协议(internet protocol)。它的功能是使主机可以把分组发往任何网络并使分组独立地传向目标(可能经由不同的网络)；定义网络中的地址方式。

• 传输层(transport layer)。它的功能是使源端和目的端主机上的对等实体可以进行会话，完成数据在应用层和互联网层之间的传输。它包括两个端到端的协议。第一个是传输控制协议TCP。它是一个面向连接的协议，把从一台机器发出的字节流无差错地发往互联网上的其他机器，并能提供排序和流量控制。第二个协议是用户数据报协议(user datagram protocol)。它是一个不可靠的无连接协议，用于不需要排序和流量控制，快速递交比准确递交更重要的应用程序。

• 应用层(application layer)。它包含所有的高层协议，最早有虚拟终端协议(TELNET)、文件传输协议(FTP)、和电子邮件协议(SMTP)。近年来又增加了不少的协议，例如域名系统服务(DNS，domain name service)用于把主机名映射到网络地址，NNTP协议用于传递新闻文章以及网络文件系统(NFS，network file system)等等。

• 主机至网络层。TCP/IP参考模型没有真正描述这一部分，只是指出主机必须使用某种协议与网络连接，以便能在其上传递分组。

3. 万维网

万维网是建立在因特网和客户/服务器模型上，以超文本传输协议(HTTP，Hypertext Transfer Protocol)、超文本标识语言(HTML，Hypertext Markup Language)和全球资源定位器(URL，Universal Resource Locator)为基础，能够提供各种Internet服务，用户界面一致的信息浏览系统。它是当前最普遍、最广泛的因特网应用平台。由于万维网能将文本、图像、声音、动画、视频等各种信息，以及新闻组、文件传输、远程登录、电子邮件

等各种服务无缝连接，因此很快就在各国政府部门、院校等机构普及开来。

从用户的角度看，万维网由庞大的、世界范围的文档集合而成，常简称为页面(page)。这些页面可放在同一主机上，也可以放在不同地理位置的不同主机上，每一页面可以包含与世界上任何地方的其他相关页面的链接，即超文本链接(Hyperlink)，万维网用 URL 来管理 Internet 上的 Hyperlink。

(1)超文本传输协议(HTTP)。标准的万维网传输协议是超文本传输协议 HTTP。它由两个相当明显的项组成：从浏览器到服务器的请求集和从另一方向来的应答集。在客户端，浏览器通过向页面所在的机器建立一个 TCP 连接，并在这个连接上发送一条消息取来所需的页面，解释它所包含的文本和格式化命令，以适当的形式在屏幕上显示。在服务器端，每个服务器站点都有一个服务器监听 TCP80 端口，看是否有从客户端过来的连接，在连接建立起来后，每当客户发出一个请求，服务器就发回一个应答，然后释放连接。定义合法的请求与应答的协议叫 HTTP 协议。如果要查看 URL 为 http://www. w3. org/hypertext/WWW/TheProject. html 的页面，其步骤如下：

- 浏览器确定 URL。
- 浏览器向 DNS 询问 www. w3. org 的 IP 地址。
- DNS 以 18. 23. 0. 23 应答。
- 浏览器和 18. 23. 0. 23 的 80 端口建立一条 TCP 连接。
- 它接着发送 Get /hypertext/WWW/TheProject. html 命令。
- www. w3. org 服务器发送 TheProject. html 文件。
- 释放 TCP 连接。
- 浏览器显示 TheProject. html 中的所有正文和图象。

(2)超文本标注语言(HTML)。Hypertext 即超文本文件。一般的文本文件(如书籍)都要求你顺序地读下去，而且版面具有格

式，超文本文件则不同，它没有版面排列格式，没有分页标记，最突出的一点是它可以通过 Hyperlink 的方式，直接跳越到你感兴趣的其他文件或页中。

万维网页面是由一个称为超文本标注语言 HTML 的语言写出的。HTML 允许用户产生包含文本、图像和指向其他万维网页面的指针(URL)的页面。

HTML 是一种标识或者说是格式化的计算机语言，通过将标准化的标注命令写在 HTML 文件中，使得任何万维网浏览器能阅读和重新格式化任何万维网页面，使之适合不同的显示器和窗口。它允许你对文本文件通过加上格式命令——一个附加指示符(directive)，放在"＜"和"＞"之间，如＜img src＝"image. jpg"＞，来格式化文字、图形、声音、动画和电影，并且可以把这些存储成标准的文本文件格式——ASCII 文件。应用 HTML，你可以做到：设计可以适应各种网络浏览器的网页；创立一些连接，可以使用户通过你的网页直接跳越到因特网上的其他网页；在因特网上传输音乐、图像、动画、电影；连接因特网上的其他资源、数据。

(3)全球资源定位器(URL)。当万维网第一次创建时，让一页面指向另一万维网页面时需要一些用于命名和定位页面的机制，URL 就是用来标识页面的世界唯一性的名字。URL 由三部分组成：协议，页面所在机器的 DNS 名字，标识指定页面(经常是一个位于它所在机器上的文件)的唯一的本地名字。例如：

http://www. cs. vu. nl/welcome. html

这个 URL 由三部分组成：协议(http)、主机的 DNS 名(www. cs. vu. nl)和文件名(welcome. html)，由特定的标点分隔各部分。

(4)公共网关接口(CGI)。CGI 是一个用于网页服务器和用户计算机之间的特定标准，CGI 允许网页用户通过网页的命令来调用一个在网页服务器的程序，并可以接受到这个程序的输出结果。在计算机领域，"网关"的定义是"可以把不同格式的信息转换成为用户可以接受的格式"的实用程序。CGI 是一个用于网页的脚本或

程序，作用类似网关，它可以接受非 HTML 格式的输入或数据，经过处理之后，输出一个标准的 HTML 格式文件，可以被各种各样的网络浏览器观看。

现在一些回答查询的万维网站点都有一个广告数据库。万维网站点会根据用户查询的内容，在数据库中取得相应的广告插入到页面中，如果用户在查询 car 时，可能会显示上海大众的广告，在查询“vacation”(假期)时，中原航空公司的广告将出现。如果用 CGI 方式来实现它，实际上就是编写一段脚本(script)或程序，作为万维网和数据库之间的接口。这个脚本被赋予一个 URL(在网页中有一个超链接指向此 URL)，位于 cgi—bin 目录下。当用户进行提交时，浏览器将和此 URL 建立 TCP 连接，由 HTTP 服务器启动该脚本程序。

开发 CGI 应用程序，可以应用许多计算机软件及语言，包括 C/C⁺⁺， Visual Basic、Perl（Practical Extraction Report Language)等。

(5)因特网程序语言。交互网程序语言是随着因特网的发展而应运而生，目前主要的因特网程序语言是由美国 SUN 公司推出的 JAVA 语言。

JAVA 语言是专门为因特网设计的一种面向对象的程序设计语言，它的最大优点，用 SUN 公司的一个口号就是“写一次，运行任何平台”。它与计算机的体系结构无关，是一种独立于操作系统的计算机语言。使用 JAVA 语言编写的程序，可以象其他程序语言编写的程序一样独立地运行，也可以插入在一个 HTML 文件中。在网络浏览器下载该 HTML 文件时，JAVA 程序的执行源文件码也同时被下载到用户端的机器上，由浏览器解释执行。这种在 HTML 中使用的 JAVA 程序，被人们称为“Applet”。

·简单：与流行的 C⁺⁺语言相比，JAVA 语言是一个既简单、又功能齐全的程序语言。它的语法结构与 C⁺⁺类似。但它没有 C⁺⁺语言中的标题文件、指针、结构、联合等复杂的类型。

·面向对象:JAVA 是一种完全面向对象的计算机语言,具有封装、多态、继承和动态联编等所有面向对象语言的特征。

·分布式:JAVA 的网络功能既强又容易。JAVA 应用软件可以很容易地通过 HTTP 或 FTP 等 TCP/IP 协议在网络上打开远程文件系统以获取有关对象的信息。

·多线程:JAVA 程序可以同时执行多种工作,而且容易管理。JAVA 具有在多处理器上操作这种并行处理的能力。

·系统独立:JAVA 编译器将 JAVA 源代码转换成一种中性的"字节码"(bytecode)而不是机器语言,这种字节码可以在任何一台 JAVA 虚拟机上运行。

·安全:由于去掉指针,JAVA 免去了由于误用或有意识操作内存引起的程序安全问题,JAVA 安全性的第二道防线是字节码的证实过程,字节码在进入 JAVA 虚拟机之前要经过严格的语法和类型检查。

§12.3 网络技术在 GIS 中的应用与实现

1. 计算机网络在 GIS 中的应用模式

网络技术在 GIS 中的应用主要体现在三个不同的层次,即局域网、广域网、万维网;从目前网络应用的模式来看,主要存在这样几种类型:

(1)主机——终端式网络。使用大型主机作为服务器,通过终端来访问数据,如图 12—3。网络传输介质使用电话线、局域网络或者专用线路。所需要的数据全部集中在主机上,事务和数据高度集中,适合于大型集中的 GIS 应用。这种系统软件专业性较强,开发难度大。系统内服务器要同时维持多个连接,负担重,对线路的要求较高。系统间只能以重复备份或脱机拷贝的方式交换数据,成本高,效率低。

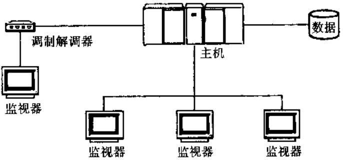


图 12－3 主机 —终端式网络

(2)工作组网络。基于工作组的对等网络，如 Windows for workgroup、Window95 等。如图 12－4。网络中的每台计算机自行管理自己的资源。通过提供文件共享来提供网络访问能力，可以适应很多种方式下的文件访问需求。很多能够通过网络运行的 GIS 都可以按照这种方式通过共享文件来共享数据资源。由于使用网络底层进行通讯，无法对安全性进行更多的控制，无法解决网络文件的访问冲突，也无法针对不同的应用进行效率的优化。这种方式的最大缺点是效率较低，很难实现广域网扩展。

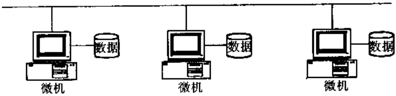


图 12－4 工作组网络

(3) 客户/服务器(C/S,Client/Server)网络。工业界把 C/S 计算模式称为"恰到好处的规模(Rightsizing)"，意指 C/S 将应用资源在客户和服务器之间进行恰到好处的分配。如图 12—5。C/S 模式把 GIS 软件的功能划分成客户单元和服务器单元，分别加载到

局域网的客户机和服务器。客户机向服务器发出数据和服务的请求，服务器对客户机的请求作出响应。通常在服务器端（称作后端系统）利用 RDBMS 进行数据的管理，处理对数据存储和访问的请求，而其他功能模块和图形用户界面则放在客户端（称作前端系统）。

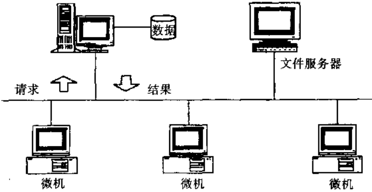


图 12－5 客户/服务器网络

虽然最初的 C/S 模式只能支持单一的操作系统和局域网，而且当客户机增加到一定数量时，服务器将被客户机的请求所淹没。但 C/S 系统的优点是明显的：通过将应用资源在前端和后端系统间的分离，降低了网络上的开销。因为在系统中，网络上传送的一般是请求（如 SQL 语句）和结果（如记录），而不是整个文件。它的另一个好处是允许多个用户同时对同一应用程序数据的访问，对数据库服务器的访问通常采用标准的 ODBC 接口，对于开发来讲也十分友好。集中式 C/S 技术已日趋成熟，几个流行的数据库厂商，如 SYBASE、ORACLE 以及 INFORMIX 等均提供解决方案，它们非常适合于中小型企业以及大型企业部门级的局域网 GIS 应用。

（4）分布式 C/S 网络。C/S 本身就是一种分布式处理技术，这里的“分布式处理系统”为了和前述 C/S 处理系统相区别，专指存

在两个或两个以上服务器的C/S系统，这种系统伸缩性更强，实现及管理也更复杂。集中式C/S系统在一个广域网或互联网环境下显然存在问题：一个部门的用户可能需要共享地理上遥远的其他部门或总部主机上的部分地理信息，而且希望"信息随手可得"，将地理空间数据在不同的主机或地点进行分布不可避免，这就导致分布式C/S系统的出现。

如图12－6。在分布式系统中，用户从本地服务器请求数据；如果本地服务器发现本地没有需要的数据，就通过LAN或WAN向其他服务器发出请求，通过"服务器对服务器的对话"处理数据，最后将数据返回给用户。整个过程中，除了一点响应延迟感觉，用户并不知道数据来自其他系统，也不必关心数据在物理上存储于何处，不管数据是在本地还是在远程。

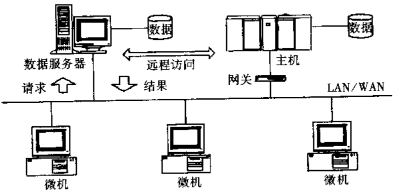
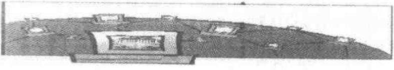


图12－6 分布C/S网络

(5)对等C/S网络。对等系统被称为C/S系统的最终归宿，是分布式处理的极限。如图12—7。C/S系统的两个不同的部分：请求服务的CLIENT和处理请求的SERVER，发展到对等服务阶段，将脱离主从结构，实现完全等同，即在网络上每个计算机既是CLIENT又是SERVER。在对等系统中每个计算机都是平等的。进一步说，每个计算机都要负责对自己资源的管理和访问并有对

任何其他计算机上的资源的同等访问权，对于对等系统而言，最好实例莫过于跨越世界的 Internet。



INTERNET　对等处理环境

图 12-7　对等 C/S 网络

近年来基于 Internet 的各种应用在全球得到了极其迅猛的增长，其成功是因为与其他技术相比，它具有开放、廉价、易于使用、易于连接并且支持多媒体数据(当然也包括地图和图形数据)等特点。对于 GIS 而言，在 Internet 上存在着巨大的应用潜力，最明显的几个方面是：

· 数据出版 在网络上提供地理图形的查询检索等服务(例如房地产公司查询不同价位的房屋的分布情况)；

· 产品销售 通过 Internet 直接向用户销售特定的地图产品；

· GIS 服务 公司可以通过 Internet 向它的客户提供在线的地址查询等服务(例如银行提供自动取款机位置查询服务)；

· 数据共享 政府部门可通过 Internet 向其上下级部门和相关机构发布本部门的 GIS 数据，从而达到数据共享的目的。

显然，Internet 为那些希望提高信誉和拓展市场的企业，同时也为提高政府部门的工作效率提供了有力的手段。在 Internet 网上应用的 GIS 也称为万维网 GIS 或 Web GIS。本节我们将重点介绍 Web GIS 实现的技术。

2. 在因特网上应用 GIS——万维网 GIS(Web GIS)

在万维网技术出现之前，因特网在 GIS 领域的应用只限于输送地理空间数据。例如使用 FTP 可以让用户从网络上把地理空间数据拷贝到用户的硬盘上，用户再用桌面 GIS 软件来处理这些地理空间数据并以地图的方式显示出来。这种方法无法直接在网上

看到地图。

利用 Telnet 和 X 视窗，用户也可以使用远程服务器上的 GIS 软件。这一方法与在一个局域网上用服务器上的软件很类似。但是这种方法需要用户在远程服务器上有帐号。而且，每个用户还必须在他的计算机上安装 X 视窗，这又受限于 Unix 操作系统。

最近对 GIS 用户及客户做了一次调查，以研究他们在 Internet 上使用 GIS 的各种需求。下面是这次调查所反映出的主要问题：

(1)使用标准工具。Internet 之所以广泛流行的主要原因之一就是因为标准的、简单易用的网络浏览器极容易得到。各种 INTERNET 浏览器(客户)，如微软的 Internet Explorer 和 Netscape 的 Navigator，以极其便宜的价钱甚至于免费就可以得到。利用它们，用户可以检索不同服务器上的数据，而不管服务器在什么地方(Web 服务器可能在同一幢建筑物中；同一个城市中或世界上的其他地方，这对用户来说是透明的)。Internet GIS 分析和操作也应该使用标准的浏览器和服务器技术，而不应该要求用户再增加非标准的工具，因为这样会增加用户的负担，并会加大出现问题的机会。更进一步，用户非常关心的是，他们的应用应该可以在异构硬件环境中并在多种浏览器和服务器上都可以运行。

(2)动态而安全的数据检索。利用 Internet 进行数据发布和 GIS 分析的一个主要原因就是它具有随时间变化而传送动态数据的潜在能力。要发挥 Internet 的最大潜力，用户更希望交互地操作数据并直接在 Internet 上出版或共享地图和其他地理信息数据。这就意味着，要提供对于一个动态地向客户提供数据服务的 GIS 数据库服务器的直接检索的能力。一种理想的方案是，利用标准浏览器通过广泛使用的数据格式对服务器进行检索，而不必进行数据转换。由于使用了 Internet 客户/服务器技术，最终用户不与数据库直接连接，使得不必花费大量资金就可以容易地维护数据库的安全性和一致性。

(3)丰富的制图和GIS分析功能。目前Internet上典型的计算模式包括在服务器上存储和处理数据，然后将结果传送到客户浏览器上。服务器上对于地理数据进行处理的能力直接影响到该系统是否能有效地为客户提供服务。目前，典型的GIS应用包括制图、设施定位(餐馆、商店、旅游点和车站等)，以及研究新区域和环境。随着INTERNET越来越遍布世界各地并广泛被大家所接受，在Internet上提供GIS服务将变得更加重要。例如，一个用户希望把他的客户地址数据传送到服务器上以把这些客户定位到地图上，再利用另外一个服务器上所提供的该地区的人口调查数据进行市场分析和规划。可以想象，在不久的将来，Internet上将出现大量专门提供特定数据和GIS服务的服务器。

(4)多种数据格式。对于用户而言，在Internet应用中，他们既希望得到栅格数据又希望得到矢量数据。栅格数据非常适合于制图应用和显示GIS处理所得到的结果。栅格形式的图由于使用区域填充和真彩色等技术，使得它们看起来更像"真正的地图"(纸图)；而另一方面，矢量数据格式对于处理和分析更加有用。看来，用矢量和栅格数据进行分析和处理，而用栅格数据进行结果的显示不失为一种好的办法。例如，ESRI公司在其INTERNET软件中就使用高度压缩的INTERNET GIF和JPEG数据格式进行结果数据的传送。

基于这样的客户需求，人们开始利用万维网技术开发万维网GIS(Web GIS)。目前Web GIS设计与实现的方法有很多，但主要有以下几种解决方案。

3. Web GIS的解决方案

(1)CGI和Web Server API解决方案。最先用于发展Web GIS技术的方法是CGI(通用网关接口)。

CGI是一种连接Web服务器上的应用软件和Web服务器的标准技术，它是HTML的功能扩展。HTML超文本是一种功能非常有限的静态文本，它的设计原则是显示数据和文件，而不是处理

数据，因而不可以用它来产生动态的信息，缺乏交互性。但是灵活易用的 CGI 程序与 HTML 的结合实现了交互式的动态通信。例如，为了让用户可以存取数据库中的信息，CGI 可以被用来连接网络服务器的数据库服务器，在系统中起通信桥梁作用。

GIS 的厂商在其成熟产品的基础上兼容发展 Internet 解决方案，通常采用 CGI 方案，即提供专用空间数据库的 Web 接口。ESRI(1997)推出了 MapObjects INTERNET/INTRANET Map Server，使用 HTML 将客户端的请求传给 Web 服务器，Web 服务器通过专用的 CGI 访问 GIS 服务器，GIS 服务器承担所有的查询、计算工作，将结果构建成一个 HTML 文档反馈给 Web 服务器，再传递给客户端浏览器。

由于 CGI 是网络服务器上的可执行程序，任何人只要可以访问相应网络服务器，就能运行 CGI 程序。Web 一出现，CGI 技术很快就被用于构造能产生动态地图的 Web 网站。基本上所有的计算机语言都可以用来开发 CGI 程序，最常用的几种包括 C/C++、Perl 和 Visual Basic。根据程序的特点，GIS CGI 技术方法可以被分为以下两类：

· CGI 启动制图软件以批处理方式运行。

一个典型的例子是美国施乐公司的地图服务器，它是由一个简单的制图软件经修改后成为以批处理方式运行的程序，它可以根据命令行参数来产生各种不同的地图。运行的结果通常是以 GIF 格式图像的方式返回用户，这种系统的长处是程序简单，运行速度比较快，但由于它的简单，比起商业化的制图软件而显得性能不足，并且对于用户的每一次请求都要重新启动制图软件。大多数的商业 GIS 软件并不是设计成以批处理方式运行的，例如，尽管 ARC/INFO 可以以批处理方式运行，但要用它作为网络服务器后端制图软件是不切实际的，因为 ARC/INFO 的启动时间较长，而且没有用户的介入是很难处理复杂地图的。

· CGI 启动后端 GIS 程序。

以前很多GIS软件是基于图形用户界面GUI用“事件”(Event)驱动的程序，一旦启动，就可以一直运行等待触发事件。这种系统有许多优点，而且网络浏览器也是利用图形用户界面，由事件驱动的，这样，一个网络浏览器上的事件，通过CGI很容易传到后端GIS软件上，CGI和后端GIS软件的信息交换是通过“进程间通信协议”IPC(Inter－process Communication)来完成的.常见的IPC协议有传统UNIX上的远程过程调用RPC，WINDOWS的动态数据交换DDE和对象连接和嵌入OLE以及MAC机上的(Apple event)，利用这种CGI方法，只要用户在网络浏览器按一下，信息就通过交互网络传到网络服务器上，然后由CGI程序将此信息通过IPC传到后端的GIS软件。例如，可以要求GIS软件将地图某个地区放大，然后将放大后的地图屏幕图像传回给用户。

这种方法的优点在于，运行速度较前一种快，因为它不需要每次启动后端的GIS软件，同时可以利用商业化GIS软件产生高质量的地图。事实上GIS软件的所有功能都可以被利用起来。然而这种方法仍有许多不足之处，首先，很难同时运行多个CGI程序，因为受软硬件的限制，可同时运行的GIS软件的份数通常是有限的，同时亦很难跟踪看出哪个用户用哪份程序。CGI结构中，CLIENT端仅起了一个哑终端的作用，其功能限于向SERVER发送用户请求和显示所接受的SERVER的处理结果，SERVER承担了一切计算功能。这种模式并没有实现真正的分布式协同计算，它仅适用于封闭环境的小型局域网，对于开放型的互联网络很容易引起服务器的过载。

其次，这种方法需要制图软件一直处于运行状态，这不仅要求GIS软件所在的服务器一直联机而且也消耗不少计算机资源。当交互网的流量较低时，这种系统会保持良好的运行状态。但多个用户同时访问服务器时，多个CGI备份同时运行会导致服务器负载过重而降低效率，使运行速度大打折扣。显然，用户产生的每一个事件都要通过交互网，由服务器来处理，当交互网流量较高时，

CGI 并不是一种理想的技术路线。其他的不足之处包括从服务器方面驱动的 CGI，必须用某个特定的数据库服务器的专用 SQL 语言手工编写数据库接口，网络安全问题以及有限的交互性问题(即用户很难直接操作地理实体)。这种系统的例子有 ESRI 的 INTERNET/INTERNET MAP SERVER (http://www. esri. com)，MAPINFO 的 PROSERVER (http://www. mapinfo. com)，Intergraph 公司 GeoMedia Web(http://www. intergraph. com)，Autodesk MapGuide(http://www. autodesk. com)。

为了解决 CGI 方法的局限，Netscape 和 Microsoft 分别推出了他们 Web 服务器的 Sever API—— NSAPI 和 ISAPI。基于 Sever API 的动态连接模块在启动后通过 IPC 与 Web 服务器进行信息交换，它会一直处于运行状态而不用重新启动，因此要比 CGI 方法快得多。不过，由于 Sever API 不象 CGI 那样有个同一的标准，因此一旦采用了某种 Sever API，那么服务器端就受限于该 Web 服务器软件和计算机平台，例如采用了 Microsoft 的 ISAPI，那么你的 Web 服务器就只能安装 Windows NT 平台和 Microsoft 的 IIS (Internet Information Server)。

(2)客户端插件(PLUG－INS)解决方案。基于 CGI/Sever API 系统仅提供给用户端(client)有限的 GIS 功能，传给用户的信息都是静态的，而且用户的 GIS 操作都需要由服务器来处理而不能在客户端进行。解决这个问题的方法是把一部分服务器上的功能移到用户端上，这样不仅加快了用户操作的反应速度，而且也减少了交互网上的流量。

标准浏览器只提供了一些最基本的浏览和导航功能，而缺乏空间数据的处理能力。一种方法是安装额外能和网络浏览器交换信息的专门 GIS 软件。这种增加网络浏览器功能的方法就叫"插入法"(Plug－ins)，为便于其他软件厂商开发插入型软件，Netscape 公司专门提供了一套应用程序接口(API)。相应的开发插件的技术还有 Java Applet 和 ActiveX。

目前这种插入软件已被普遍采用，在多媒体领域尤为明显。这种插入软件不但可以增加网络浏览器处理地理空间数据的能力。使人们更容易获取地理数据，而且可以减少网络服务器的信息流量从而使服务器更有效地为更多的用户服务，因为大多数用户的数据处理功能可以由网络浏览器插入软件来完成。例如 Autodesk 为 INTERNET/INTRANET 提供了一种新的二维矢量文件格式 DWF 和称为 WHIP！Plug—in 的浏览工具，以便用户能有效地从 INTERNET/INTRANET 下载工程技术及设计类的数据资料，借助这些新技术，用户可以查阅二维矢量数据，并可利用高速缩放和平移的功能迅捷的浏览内容复杂而详尽的图形文件。此外，对嵌入的 URL 的支持可允许用户在 INTERNET/INTRANET 中与其他地点建立直接的联系，从而使世界各地的设计人员能够通过 INTERNET/INTRANET 在设计项目上进行方便的合作，以便与客户分享他们的工作结果。Autodesk 针对设计和工程技术领域而创建的这项技术，有助于减少工程师、建筑师和其他设计人员之间进行交流的障碍，并使他们能更多地共享数字化设计信息。DWF 文件格式比位图或其他存储和显示设计信息的文件格式更有效。位图文件格式如 GIF 格式和 JPEG 格式都以单个像素来存储数据，而矢量格式却用直线、圆弧和圆来存储图形。为了进一步减少装卸文件的次数，DWF 文件以压缩形式传递，然后用 WHIP！Plug—in 复原。由于设计文件的规模经常具有几兆字节，因而这种压缩技术就成为影响设计效率的关键因素。浏览图形时采用的缩放和平移操作几乎是实时完成的。因为这不需要重装图形或访问服务器。

尽管插件可以和浏览器一起有效处理空间数据，但这种方法仍有缺陷。首先，它将导致用户端负担过重(Fat Client)，因为几乎每个软件厂商都希望它的软件能与交互网兼容(INTERNET/INTRANET—Enabled)。这显然不符合标准网络浏览器方便简单(Thin Client)的设计思想。其次，众多的插入软件的管理会成为信

息技术部门的一个大问题，因为任何人只要可以连上交互网都可以接收最新的插入软件。和传统的应用软件类似，插入软件也需要先安装再使用，因而传统软件不同版本之间的不兼容性及版本管理问题也仍然存在。

(3)JAVA Web GIS 解决方案。为解决上述各种问题，交互网程序语言应运而生。交互网程序语言的出现标志着 Web GIS 的开始，目前最普及的交互网程序语言是由 SUN 公司开发的 Java，尽管出现时间较短，但发展迅速，几乎所有重要软件公司都在研制与 Java 兼容或直接用 Java 开发的产品，且许多已经上市。Java 是一种面向对象的计算机语言，它借鉴了 C、Smalltalk、Object C^{-+} 和 Cedar/Mesa 等面向对象语言的优点，其特点是简洁、动态适应性强，运行稳定、安全，对网络而言与计算机结构体系无关，容易移植，在一种系统下开发的应用软件可以直接在完全不同的系统下运行。事实上 Java 编译器产生的是一种独立于任何操作系统的字节码 Bitecode，这种字节码程序可以在任何一台 Java 虚拟机(Java Virtual Machine－JVM)上运行，任何系统只要支持 Java 虚拟机就可以运行 Java 程序，而与程序在何种系统下开发和编译无关，目前 Netscape 和微软公司的网络浏览器都直接支持 Java 程序。

第二种交互网程序语言是由微软公司提出的 Active X，其实 Active X 控制是由 OLE 控制 OCX 加上一些新的交互网的界面函数发展起来的，Active X 仍然依赖现有 OLE 编程体系来达到增加交互网交互性的目的，方法之一是利用一个 OLE 文件实体 DocObject 作为一个通用控制容器 Container，例如，微软公司的网络浏览器 Explorer 就可以发展成为一个文件实体的控制容器，然后加入 GIS 引擎作为文件实体的服务器函数。这样扩充后的网络浏览器就能显示和处理地理空间数据。Active X 也可以作为服务器，例如，微软公司的分布式 COM(Distributed Common Object Model)，将使在用户和服务器两端的 Active X 控制互相交换信息，从而把整个网络上的负荷分布到各个不同的子网上。与 Java

相比，Active X 目前还没有解决非常重要的网络安全问题。

交互网程序语言与前面所述方法的主要不同之处在于允许可执行程序从网络服务器上传到用户机器上，并直接在用户机器上执行，从而省略了一整套传统软件的安装过程。比较而言，这种方法有许多优越之处，容易分发，不要安装且可以集中版本控制。由于程序是在用户端执行的，从而避免了用户和服务器之间不必要的信息流量。它在用户端运行时就象通常的软件一样，当需要数据时，客户机对象可以直接调用服务器对象，通过交互网从服务器获取数据，而不需要 HTTP 和 Web 服务器作为中介。这样不仅减少了网络信息流量，而且也提高了整个网络的运行效率，再者它也提高了整个万维网 GIS 的效率。比如图形放大缩小这种常用的操作，就可以在用户端的机器上执行，反应就会快得多。如果用前述的 CGI 技术方法，每一次地图的放大、缩小都必须由远程服务器来处理，这个交互过程在用户增多时会很慢。这种技术方法让用户可以自由地处理每个地理实体，而不只获取由服务器处理好的静态图像。将 JAVA 作为实现 Web GIS 的主要技术，因为：

· 浏览器的发展趋势是执行网络应用程序的操作系统，JAVA Applets 从服务器下载后可在客户端运行而服务器只有 I/O 开销。

· Java 本身是一门适合于分布式计算环境的网络开发工具，并且通过 JDBC/JAVA SQL 实现了与关系型数据库的连接。

4. 几种主要 GIS 软件的网络化解决方案

由于网络软硬件的进步和用户需求的不断增大，GIS 软件的网络化趋势越来越强烈，一些著名的 GIS 厂商都相继推出可以网络化运行的软件版本。

(1)ESRI。ARC/INFO 可以将其作为服务器来运行，前端使用 ArcView 可以通过文件共享方式访问服务器数据，或者通过 SDE 来访问数据服务器的数据，从而提供企业级的网络化 GIS，可以使用 Map Objects 来进行开发。在 Internet 上使用 Map

Server 来提供服务，前端使用图形浏览控件显示图形或者直接返回图像文件。

（2）Intergraph。其著名的 MGE 模块化 GIS 软件可以在局域网上运行，图形数据通过文件共享方式提供网络存取，属性数据使用统一的 DBMS 来存储。前端可以使用 VistaMap 实现地图与属性数据的联合查询。新推出的 GeoMedia 面向网络设计，可以通过局域网访问包括 MGE、Arc/Info、SDO 等多种不同的编程接口，大大方便了开发人员，有利于开发出完全用户化的功能强大的应用软件，并且能与 MIS 系统实现无缝集成。同一系列的 GeoMedia Web Map 通过使用客户端插件和 ActiveCGM 图形格式来实现面向 Internet 的地图数据发布。

（3）Bently。其制图软件 MicroStation 可以在局域网上运行，以文件方式共享数据。新推出的 GeoGraphics 可以通过 ODBC 实现更加方便的属性数据的管理和访问。该公司用于网络数据出版的 ModelServer/Discovery 可以动态出版图形数据，并可以实现属性数据的查询。其出版的文件格式较多，包括 SVF、CGM、VRML 等，客户端使用相应的插件来显示。据称在未来版本中，将提供纯 Java 的前端显示方式。服务器方通过运行多个 MicroStation 实例来提供数据服务。

（4）MapInfo。MapInfo 可以在局域网上运行，以文件方式可以共享图形数据，可以 ODBC 方式访问属性数据库。其 Spatial Ware 可以使用 Oracle 等大型数据库来管理图形数据。MapX 是用于用户化应用开发的控件，在以后的版本中将可以直接与 SpatialWare 进行交互。ProServer 是 MapInfo 公司用于在 Internet 上出版地图数据的服务软件，主要通过调度多个 MapInfo 实例来提供数据出版与转换服务。

综合目前 GIS 软件厂商实现网络化的不同方案，虽然各有不同，但也有一些共同的特点：

• 在局域网上主要通过文件方式来共享数据。但可以看出一

个明显的趋势是服务器部分都逐渐向数据库服务器方向转移。GeoMedia 的组件模型有可能提供分布式对象服务。

• 面向 Internet 的数据出版后台服务器几乎大部分采用了原有的软件附加一调度和监控程序来完成，而不是完全开发一个服务器程序。

• Internet 网络数据存取主要使用 CGI 方式(ESRI、Bently 等公司使用了 Server API)。由于这种方式的效率不高，因而限制了其应用范围。

下面我们详细介绍两种当前比较流行也比较成熟的 WebGIS 产品：ESRI 的 IMS (Internet Map Server) 和 MapInfo 的 ProServer。

(5)MapInfo ProServer 在 Internet/Intranet 中实现地图应用。MapInfo ProSever 是专为 Internet/ Intranet 体系结构的信息系统设计的。ProSever 可实现 Professional 的全部功能，并通过网络共享。信息管理员只需在 Web 服务器上安装 ProSever，并对其进行编程和管理，用户可通过网络浏览器访问 ProSever，并获得 ProSever 上程序所提供的功能，如显示地图、访问地图上连接的信息，制作专题图，进行地理分析等，ProSever 的开发可使用 MapBasic 结合各种流行开发工具，VB、Delphi、VC++等。

使用 ProSever，开发人员能集中地控制和维护地图和数据库数据，并集中实现应用程序功能，避免了系统维护同步困难的问题。公司能够在 Sever 端统一保存和维护数据和软件，通过运行多个 MapInfo 实例，减少了在每个 Client 端购买多个单机版的投资，大大降低用户的使用(调用数据、更新数据和软件)费用。

图 12—8 是 ProSever 的体系结构。

MapInfo ProServer 使 MapInfo Professional 能在 Server 上运行，通过网络回答用户的要求。在 Client 端可以通过任何一个 Web 浏览器直接使用功能强大的 Professional。

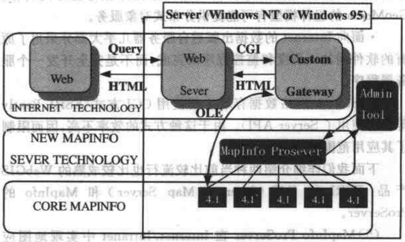


图 12－8 **MopInfo ProSerrer 的体系结构**

MapInfo ProServer 需要与一个开发人员定制的应用软件共同工作，该应用软件可以管理用户的 Internet 通用网关接口(CGI)与多个 MapInfo Professional 实例之间的交互，它还包括 MapInfo ProServer 管理器，系统管理者能够设置参数来运行 MapInfo Professional。

(6)IMS。

①IMS 的基本技术：

—— 动态 HTML(HyperText Markup Language)

对于静态 HTML，只要在服务器上定义好 HTML 主页，任何人都可以通过网络获得此服务器的信息资源，就象浏览一本书一样浏览远程服务器上的数据。但对 GIS 应用来说，用户往往按照自己的兴趣从服务器上获取动态的图形数据和属性数据，静态的主页不能满足要求，需要采取类似于双向交互式方法的动态主页来实现——根据用户的指定从数据库里查询数据生成主页，分发给用户。

—— CGI(Common Gateway Interface)

CGI 是网络上的一种技术，它是 Web 服务器调用外部程序的接口，通过 CGI 技术可在 Web 服务器上运行一个副本用来接收用户请求。并可利用 Web Server 动态分发数据。Web 客户端通过浏览器发送一个请求到 Web 服务器，服务器通过 CGI 把请求转发给后端运行的 GIS 应用程序，GIS 应用程序把生成结果交还 Web 服务器，Web 服务器再把结果传递给用户端的浏览器显示。

——OLE(Object Link and Embeded)

——GIS

②IMS 组成：

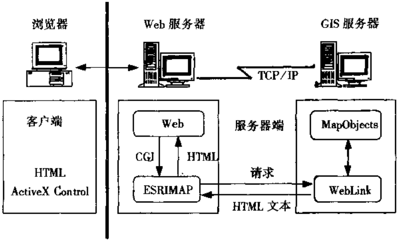


图 12—9 IMS 的组成

——Internet IMS 扩展 ESRIMAP. DLL

具体在 IMS 里，Web Server 上运行一个 IMS 扩展的 DLL (ESRIMAP. DLL，这是一个 CGI 应用)接收来自浏览器的命令，ESRIMAP 生成一个 HTML 交给 Web Server，由 Web Server 分发给 Client，这就是 IMS 里实现 HTML 的技术路径。

ESRIMAP 在 IMS 中只是一个与 Web Server 的联络者和其请求的分发者，要响应用户需求必须由另外的模块来实现，这种把

任务的分发和处理独立分开是为了支持多服务器结构，即 Web Sever 和 GIS 应用可放在不同的服务器上。一个 Web Sever 可支持多个 GIS 服务器，同样，一个 GIS 服务器可响应多个 Web Server 的请求。

——MapObjects 端 WebLink 扩展

WebLink 放在 GIS 服务器上，通过 TCP/IP 协议与 ESRIMAP 通讯，接收来自 Web Server 的各种请求，把任务交给 MapObjects 完成，并把结果以 HTML、图像、文本的形式返回 Web Server，由 Web Server 分发给用户。

—MapObjects

MapObjects 为开发者提供了一系列 GIS 和制图功能，是 IMS 中图形和属性数据的生产者。它包括一个 ActiveX 控件和 35 个以上可编程的 Active Automation 组件。

——大量用来开发客户端界面的 ActiveX 和 Java Applets 程序例子

③IMS 工作原理。服务器端有两个服务器 Web Server 和 GIS Server，这两者是通过 TCP/IP 协议进行通讯的，所以即可以运行在一台主机上，也可分别运行在两台计算机上。

在 Web Server 端运行了一个扩展 DLL：ESRIMAP，在 GIS Server 端运行 WebLink 应用，通过这两部分有效的把 Web Server 和 GIS Server 连接起来。

在 GIS Server 端，WebLink 通过 MapObjects OLE 获得强大的制图功能，包括与 ESRI 的 SDE 连接。

主要参考文献

[1]黄杏元,汤勤.地理信息系统概论.北京:高等教育出版社,1989

[2]兰运超等.地理信息系统原理.广州:广东地图出版社,1991

[3]尹贡白等.地图概论.北京:测绘出版社,1991

[4]彭望禄.遥感数据的计算机处理与地理信息系统.北京:测绘出版社,1991

[5]龚健雅.整体SIS的数据组织与处理方法.武汉:武汉测绘科技大学出版社,1991

[6]王家耀,邹建华.地图制图数据处理的模型方法.北京:解放军出版社,1992

[7]徐庆荣,杜道生等.计算机地图制图原理.武汉:武汉测绘科技大学出版社,1993

[8]刘勇卫,贺雪鸿译.遥感精解.北京:测绘出版社,1993

[9]陈耀西等.机助军事地形分析.北京:解放军出版社,1993

[10]李德仁等.地理信息系统导论.北京:测绘出版社,1993

[11]周成虎等.地理信息系统概要.北京:中国科学技术出版社,1993

[12]邬伦等.地理信息系统教程.北京:北京大学出版社,1994

[13]张海藩.软件工程的概念与方法.北京:中国科学技术出版社,1994

[14]马建文,阎积惠.地理信息系统及资源信息综合.北京:地质出版社,1994

[15]华一新．计算机专题地图制图．郑州：解放军测绘学院教材，1994

[16]韩丽斌．地图数据库原理与技术．郑州：解放军测绘学院教材，1994

[17]面向对象的空间数据库译文集．西安：总参测绘研究所，1994

[18]杜道生，陈军等．RS、GIS、GPS 的集成与应用．北京：测绘出版社，1995

[19]刘鲁．信息系统设计原理与应用．北京：北京航空航天大学出版社，1995

[20]黄幼平等．GIS 空间数据误差分析和处理．武汉：中国地质大学出版社，1995

[21]孙家广等．计算机图形学．北京：清华大学出版社，1995

[22]梁启章．GIS 和计算机制图．北京：科学出版社，1995

[23]中国地理信息系统协会首届年会论文集．北京：中国地理信息系统协会，1995

[24]95'3S 技术研讨会论文集．北京：中国测绘学会等，1995

[25]边馥苓．地理信息系统原理和方法．北京：测绘出版社，1996

[26]中国地理信息系统协会第二届年会论文集．北京：中国地理信息系统协会，1996

[27]林珲等．城市地理信息系统研究与实践．上海：上海科学技术出版社，1996

[28]马智民等．应用地理信息系统设计与实现．西安：西安地图出版社，1996

[29]地理信息系统软件工程及相关技术高级研讨会论文集．武汉：武汉测绘科技大学，1997

[30]华一新．地理信息系统原理与技术．郑州：解放军测绘学院教材，1997

[31]王家耀,华一新.军事地理信息系统.北京:解放军出版社,1997

[32]史文中.空间数据误差处理的理论与方法.北京:科学出版社,1998

[33]王家耀,徐国华,华一新.数字制图技术与数字地图生产.西安:地图出版社,1998

[34]Singh 等.Oracle 数据库开发指南.北京:清华大学出版社,1998

[35]张莉,周伯庄等.标准建模语言 UML 及其支持环境.北京:计算机世界,1998

[36]陈俊,宫鹏.实用地理信息系统.北京:科学出版社,1998

[37]高俊等.虚拟现实在地形环境仿真中的应用.北京:解放军出版社,1999

[38]修文群,池天河.城市地理信息系统(GIS).北京:北京希望电子出版社,1999

[39]Roger S. Pressman.软件工程——实践者的研究方法.北京:机械工业出版社,1999

[40]中国 21 世纪议程管理中心.中国地理信息元数据标准研究.北京:科学出版社,1999

[41]Ralph M. Stair,George W. Reynolds.信息系统原理.北京:机械工业出版社,2000

[42]陈述彭等.地理信息系统导论.北京:科学出版社,2000

[43]华一新.专题地图设计专家系统 PC—MAPPER(博士论文).郑州:解放军测绘学院,1991

[44]陈荣国.面向对象图像信息系统的设计原理、方法和实践(博士论文).郑州:解放军测绘学院,1993

[45]徐青.三维可视化技术的研究与实践(博士论文).郑州:解放军测绘学院,1995

[46]白玲.GIS 中地理网络分析的研究与实践(博士论文).郑

州:解放军测绘学院,1997

[47]蔡少华.GIS 图形空间关系的研究与实践(博士论文).郑州:解放军测绘学院,1999

[48]M. F. Goodchild & K. K. Kemp. Introduction to GIS. National Center for Geographic Information and Analysis University of California,1991

[49]G. Gambosi M. scholl H. — W. Six(Eds.). Geographic Database Management Systems. Workshop proceedings, Capri, Italy,May,1991

[50] Stewart Fotheringham & Prter Rogerson. Spatial analysis and GIS. Taylor & Francis,1994

[51] Stan Aronoff. Geographic Information Systems: A Management Perspective. WDL publications Ottawa ,Canada, 1991

[52]M. F. Wroboys. GIS A Computing Perspective. Taylor & Framcis,1995

[53] Philippe Krechten. The RATIONAL UNIFIED PROCESS: An Introduction . Addison Wesley Longman, Inc, 1998

[54]Ivar Jacobson, Grady Booch, James Rumbaugh. The Unified Software Development Process. Addison Wesley Longman,Inc,1999

(本书由总参谋部测绘局组织编写)